

THE HUMAN, LOVE, AND THE INNER LIFE: ETHICS AFTER MURDOCH

by

Annalisa Paese

Laurea Triennale (First Degree) in Philosophy, Sapienza Università di Roma, 2002

Laurea Magistrale (Post-Graduate Degree) in Philosophy, Sapienza Università di Roma, 2004

Submitted to the Graduate Faculty of the
Dietrich School of Arts and Sciences in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH
DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Annalisa Paese

It was defended on

July 19th, 2018

and approved by

John McDowell, Distinguished Professor of Philosophy, Philosophy Department

Michael Thompson, Professor, Philosophy Department

Jennifer Whiting, Distinguished Professor, Philosophy Department

Edouard Machery: Professor, Department of History and Philosophy of Science

Dissertation Advisor:

John McDowell, Distinguished Professor of Philosophy, Philosophy Department

Copyright © by Annalisa Paese

2018

THE HUMAN, LOVE, AND THE INNER LIFE: ETHICS AFTER MURDOCH

Annalisa Paese, PhD

University of Pittsburgh, 2018

Moral philosophy, it is often pointed out, can easily end up demolishing rather than advancing our ethical self-understanding. I offer an interpretation of Iris Murdoch's analysis of this phenomenon that bears on questions about the moral status of human beings, the proper objects of moral assessment, and the relation between morality and love. According to the view I develop, moral philosophy is today, as it was at the time Murdoch was writing, largely unable to countenance the importance that the *concept of a human being*, *inner life*, and the *capacity for love* have in our ethical self-understanding. This is because of the enduring prevalence of the two assumptions Murdoch identified as the sources of this inability. These assumptions are (1) that volition and cognition are sharply separated powers, and (2) that empirical sciences set the standards for factuality. I argue that recognizing that these assumptions are not rationally compulsory, thereby freeing moral thought of their influence, makes possible a better understanding of the humanities and of their relation with the sciences.

TABLE OF CONTENTS

1.0	INTRODUCTION: THE HUMANISTIC AND THE SCIENCE-BASED STUDY OF HUMAN LIFE	1
1.1	POSITIONS, PICTURES OF THE SOUL, AND MORAL THEORIES	3
1.2	THE PREVAILING PICTURE	7
1.2.1	Wittgenstein and the Keystone of the Prevailing Picture	8
1.2.2	Kantian Liberalism as a Misunderstanding of Freedom	12
1.2.3	Uses of Freud and the Failure to Take the Historical Individual into Account	16
1.3	THE COGNITIVE AND THE PRACTICAL	20
1.3.1	Understanding Choice and Freedom	21
1.3.2	A Contrast Case: Anscombe's Modern Moral Philosophy	24
2.0	THE CONCEPT OF HUMAN BEING IN MORAL THOUGHT	32
2.1	A CRITICISM OF THE MORAL USE OF THE CONCEPT OF A HUMAN BEING	34
2.1.1	A Familiar Concept Under Scrutiny	34
2.1.2	Manne's critique of Humanism	39
2.1.3	A Line of Defense for Humanism	48
2.2	A DIFFERENT LINE OF DEFENSE	53
2.2.1	Dehumanization is Independent of Hostility	53
2.2.2	Explanation in Moral Psychology	57
2.3	A DIFFERENT LINE OF DEFENSE FOR HUMANISM	61

2.3.1	Recognition.....	61
2.3.2	The Practical Value of the Concept of a Human Being	64
2.3.3	Summary	68
3.0	MURDOCH AND PRIVACY.....	71
3.1	MURDOCH’S CRITICISM OF THE GENETIC ANALYSIS	71
3.1.1	Privacy and Moral Progress	71
3.1.2	Two Interpretative Difficulties	73
3.1.3	Is Murdoch’s Conception of Privacy Mythical?	79
3.2	IDIOSYNCRASY AND THE IDEA OF PERFECTION.....	85
3.2.1	Idiosyncrasy.....	85
3.2.2	Imagination and Fantasy	90
4.0	MURDOCH AND LOVE.....	95
4.1	TEMPERAMENTS AND STRATEGIES	95
4.2	TWO UNSATISFACTORY READINGS OF MURDOCHEAN LOVE.....	99
4.2.1	“Be ye therefore perfect”: the impartiality of Murdochean love	99
4.2.2	The “fabric of being”: Murdoch’s integrated picture of practice and cognition.....	108
5.0	CONCLUSION	118
	BIBLIOGRAPHY	121

1.0 INTRODUCTION: THE HUMANISTIC AND THE SCIENCE-BASED STUDY OF HUMAN LIFE

This dissertation addresses a problem posed by the relation between two ways of thinking about the practical situation of human beings. The very claim that there are two significantly different ways to approach the understanding of human life in its practical dimension is non-trivial, and part of my aim is bringing into view what I take to be a fruitful understanding of this contrast. In a preliminary fashion, though, we can think of them, respectively, as the *humanistic* and the *science-based* study of human life.

The problem arises when it is observed that human beings are just natural beings and, from this, it is inferred that the study of them belongs to the domain of the empirical sciences. Claims about human life in its practical dimension, now thought of as belonging to a region of this domain, are accordingly interpreted as the sort of claims that require a justification coming from research in these fields. When this kind of justification is missing they are treated as suspicious, if not positively fraudulent.

The general idea is that the only genuinely cognitive way of dealing with nature—and, therefore, with human beings who are a part of it—is the one distinctive of the empirical sciences. It follows from this idea that claims about human beings, if they purport to be cognitive, belong either to empirical science or to its more rudimentary and less rigorous version: folk science. Either

way, they are subject to scientific standards of adequacy, and it is, therefore, for empirical research to vindicate or debunk them.

The application of scientific standards in the interpretation and assessment of claims about human life put forward in what I am going to call humanistic contexts, though, strikes us as somehow inadequate. And this is not because, in the relevant cases, the results of the scientific investigation of human beings allegedly impugn those claims. Even when these results are purportedly offered in support of pieces of humanistic self-understanding, the very idea of treating the latter as scientific hypotheses strikes the unprejudiced mind as distorting them. Expressions of what we know, or think we know, about human life in virtue of the fact that we are living one seem to become something different once they are reinterpreted to fit into a format that allows for empirical testing. My aim is to provide an understanding of what this difference consists in.

This understanding takes the form of an articulation of the specific *perspective* and *point* of humanistic thought on human life through a contrast with those proper to scientific thought on the same subject. Reaching a positive characterization of the features of humanistic thought on human life, however, is the overarching goal of the dissertation; in this introduction, I am going to characterize them only negatively. The perspective of humanistic thought on human beings is not that of an observer, and its point is not to predict and manipulate their conduct, conceived of as the measurable effect of replicable initial conditions.

The idea that there is a kind of knowledge of human life which is different from what we pursue through empirical sciences is implicit in the writings of many philosophers in virtue of the very fact that such writings exemplify claims to that knowledge. In this dissertation, I will focus on Iris Murdoch's understanding of this kind of knowledge since, as noted by Cora Diamond, she not only aims at contributing to it, she also makes of its possibility and of the method which is

proper to its pursuit objects of direct philosophical attention (Diamond 2011). I think that reading Murdoch's work in the light of this general understanding of its significance serves to remind us of genuine possibilities for practical philosophy that tend to be overlooked. My starting point in defending this view will be Murdoch's criticism of contemporary moral philosophy in *The Sovereignty of Good*.

1.1 POSITIONS, PICTURES OF THE SOUL, AND MORAL THEORIES

In order to understand Murdoch's criticism of contemporary moral philosophy, it is helpful to start by identifying two different targets, and we may do so by looking at the way in which she describes her project. A characteristically difficult but illuminating formulation of Murdoch's understanding of the aim of her work in moral philosophy is to be found at the beginning of "The Idea of Perfection", the first of the three essays collected in *The Sovereignty of Good*. There she does not identify her target primarily in terms of a theory that she takes to be false and therefore to be refuted, but rather in terms of a "position" in which she thinks we, as moral thinkers, have come to find ourselves and from which we need to move away. What she finds unsatisfactory in this position is that "it ignores certain facts and at the same time imposes a single theory which admits of no communication or escape into rival theories" (Murdoch 1970: 1).

This claim might strike one as puzzling in many ways. It is far from evident that any *one* theory is in fact imposed by "current moral philosophy"—whether we construe this expression broadly, as post-Kantian moral philosophy, or narrowly, as moral philosophy in the decades during which Murdoch was writing. Apart from this, what is it exactly for a theory "to admit of communication or escape into rival theories"? The most straightforward way of interpreting this

phrase does not refer to something that is obviously required or even desirable for a theory. A theory might identify the problems of a certain area of thought differently from its rivals and, in this sort of case, it can be said not to admit of communication with them. This however does not seem to be a flaw in so far as the theory's way of identifying the problems is compelling. Furthermore, theories are supposed to be generally compelling, i.e. with respect to the identification of their own aims and otherwise; and they are said to be so precisely when they strike us as something we *must* think—when, in this sense, they leave no escape into rival theories. Despite the difficulties it raises, I think there is much to be gained by taking seriously Murdoch's presentation of her objectives.

When Murdoch talks about the theory for which she wants to provide a rival, she is not referring to a particular moral theory in the ordinary sense, but rather to a very general representation of the human condition—or, as she puts it, to a “picture of the soul” (Murdoch 1970: 2)¹—which, in her view, functions as the background shared by thinkers generally considered very different from each other. The provocative claim that contemporary moral philosophy is dominated by one single theory is the expression of Murdoch's conviction that, in each of the purportedly distinct contemporary moral theories, the substantive work is done by the picture of the soul they all take for granted, so that they count, at bottom, as variants of the same theory.²

Though Murdoch argues at length against this theory or picture of the soul, which I will refer to as *the Prevailing Picture*, her main goal is to challenge the assumptions that make it seem compulsory. This is, I think, the significance of Murdoch's claim that what she finds unsatisfactory

¹ She refers to this target of her criticism also as a “type of moral psychology” (1970: 4), a “theory of inner life” (1970: 9), and an “image of personality” (1970: 15).

² According to Murdoch there is really no distance between the picture of the soul and the theory. This is why she characterizes both as “existentialist”.

is the *position* that imposes a single theory—even if she clearly finds that theory unsatisfactory as well.

One can recognize, therefore, two strands in the sustained criticism that Murdoch develops throughout the three essays in *The Sovereignty of Good*. The first aims to show a kind of project in philosophy of mind or philosophy of language that purports to be a morally neutral analysis for what it really is, namely, a picture of the individual that, far from being morally neutral, is informed by specific moral values and circularly reaffirms them. The second aims to assess this picture in the way that is appropriate for it. Deprived of its appearance of being a mere consequence of scientific and logical truths, the Prevailing Picture has to face critical comparison with other similarly circular and non-neutral pictures and prove itself by its comparative capacity to enhance our understanding of moral phenomena and provide orientation for our conduct (Murdoch 1970: 44).

The arguments Murdoch uses in pursuing these two aims are tightly connected,³ but there is an approximate division of labor between two kinds of arguments. Murdoch attacks the point of view that makes the Prevailing Picture seem compulsory by objecting to certain uses of Wittgenstein's philosophy and to what she takes to be "inexact ideas of science which haunt philosophers and other thinkers" (Murdoch 1970: 26). The Prevailing Picture itself is criticized by arguments in moral psychology, involving a distinctive kind of appeal to experience that marks Murdoch's peculiar variety of empiricism.⁴ With respect to both these kinds of arguments, the

³ Cf. Murdoch 1970, p. 9: "I have simple empirical objections (I do not think people or essentially 'like that'), I have philosophical objections (I do not find the arguments convincing), and I have moral objections (I do not think people *ought* to picture themselves in this way). It is a delicate and tricky matter to keep these kinds of objections separate in one's mind."

⁴ On this topic, see Diamond 2010 and 2011.

most important move Murdoch makes involves appealing to the role of the idea of perfection in our moral life.

We can now address the question about what Murdoch means when she says that contemporary moral philosophy imposes on us a theory that does not admit of communication or escape into other theories. The problem is that this kind of philosophy constitutes a perspective from which the Prevailing Picture misleadingly appears to be a neutral and unquestionable representation of what simply must be the case because of science and because of logic. As a result, a whole range of different pictures of the soul are not taken into account but rather re-described in the terms of the Prevailing Picture, so that their characteristic values, having become inexpressible, are replaced by those characteristic of the Prevailing Picture.⁵ The values built in in the Prevailing Picture, then, are not exposed to critical comparison with possible rivals, and it is not the compellingness of this picture that blocks the possibility of communication or escape into other theories, but the fact that it is thought to have the status of “ultimate” theory in virtue of interconnected and misguided ideas about logic and science.

Murdoch works at dismantling these ideas and seeks to move the comparison between the Prevailing Picture and the rival picture she proposes on to what she takes to be its proper terrain, which is not their respective standing vis-à-vis logic and science, but the assessment of their respective power as resources for moral understanding and moral progress. While this comparison shows the inadequacy of the Prevailing Picture, the outcome is not an exclusion of its central value, i.e. freedom, but a compelling argument to the effect that we need a different way of understanding

⁵ The idea that the purportedly neutral analysis of contemporary philosophy eliminates the specificity of different moral outlooks in order to make them fit a single model is articulated by Murdoch already in “Vision and Choice in Morality;” e.g. Murdoch 1956: 45, 46-47 and 51-52. For this idea see also Murdoch 1998: 71.

it. This fact, according to Murdoch, marks a structural difference between her proposed framework and her target and makes the former explanatorily superior to the latter.

1.2 THE PREVAILING PICTURE

What, then, according to Murdoch, is the picture of the soul (i.e. the moral psychology) that dominates contemporary moral philosophy? Murdoch presents it as an image held in place with the contribution of many different forces; she describes it as “a happy and fruitful marriage of Kantian liberalism with Wittgensteinian logic solemnized by Freud” (Murdoch 1970: 9). This characterization of the picture that Murdoch wants to attack is striking in the light of the fact that she credits each of the figures mentioned in it with extremely important insights, which are integrated into her own picture of the soul. As she often emphasizes, however, her criticism of the Prevailing Picture is to be read as addressed not to doctrines that she ascribes to Kant, Wittgenstein, and Freud, but to a particular way of appropriating and combining elements from each of these thinkers’ views.

The core of the Prevailing Picture consists of two connected ideas. The first is a conception of the world as a mechanism whose workings are fixed by science and logic and as therefore inadequate to countenance anything like a moral agent or a moral reality. The second is a correlative conception of the individual as a center of pure will isolated from the mechanism. This pair of ideas is clearly recognizable as the central feature of the modern understanding of our relation with nature: the arriving point of a revolution that culminated with Kant. With an unconventionally inclusive use of the term, Murdoch indicates in these two connected ideas what makes the dominant conception of human personality *existentialist*.

In the rest of §1.2 I will explain Murdoch's view that Kant, Wittgenstein, and Freud have been read as providing resources to justify and develop this kind of moral psychology even though they didn't themselves subscribe to it.⁶

1.2.1 Wittgenstein and the Keystone of the Prevailing Picture

Murdoch thinks that the most important aspect of the Prevailing Picture is its lack of any serious consideration of the inner life, i.e. its idea that inner life is irrelevant and possibly non-existent. It is for this reason that she thinks that the argument on which the entire framework depends (she calls it its "keystone") is an argument that purports to show that mental concepts, contrary to our ordinary understanding of them, are not connected to the inner life of the subject. The upshot of this crucial argument, which Murdoch refers to also as "the genetic argument," is that the meaning of mental concepts is exhausted by the publicly observable circumstances on which one relies in acquiring them, and therefore the inner occurrences that we pre-theoretically tend to associate with

⁶ Even if Kant is widely and authoritatively read as the thinker who provides the fullest elaboration of what Murdoch calls existentialism, she thinks that there are elements in his view that set his moral psychology apart from the one that constitutes her main target.

Christine Korsgaard provides an example of the reading of Kant as an existentialist in Murdoch's sense: "[F]or us, the world is no longer first and foremost form. It is matter. This is what I mean when I say that there has been a revolution, and that the world has been turned inside out. The real is no longer the good. For us, reality is something *hard*, something which resists reason and value, something which is recalcitrant to form. If the real and the good are no longer one, value must find its way into the world somehow. Form must be imposed on the world of matter. This is the work of art, the work of obligation, and it brings us back to Kant. And this is what we should expect. For it was Kant who completed the revolution when he said that reason—which is form—isn't in the world, but is something that we impose upon it" (Korsgaard 1996: 5).

For Murdoch's resistance to this kind of reading of Kant, see for instance Murdoch 1970: 30.

these concepts have no existence independently of such observable circumstances (Murdoch 1970: 9-10, 12).

The genetic argument, according to Murdoch, is a development of a more general one, which, unlike the development itself, is in fact to be found in Wittgenstein's discussion of privacy. This argument is more general than the target of Murdoch's criticism in that it concerns both what she calls "mental" concepts, such as "repentance", and what she calls "simple, non-mental" concepts, such as "red". Both the former and the latter, Murdoch argues, had been thought by philosophers to derive their meaning from inner objects conceived of as exclusively and infallibly accessible to the subject. Wittgenstein, as Murdoch reads him, showed that both this understanding of meaning and the very notion of inner objects it appeals to are confused. However, this observation does not warrant the more specific conclusions about mental concepts mentioned above, that is, that they have to be analyzed genetically, so that the inner is to be understood as a mere possible accompaniment of what truly constitute their instantiations, i.e., the obtaining of publicly observable facts. That these conclusions are inadequate, in her view, is particularly evident and salient when we consider certain interesting uses which mental concepts characteristically lend themselves to and which she calls *specialized personal uses* (Murdoch 1970: 25).

In Chapter 3 I will address in more detail Murdoch's idea of how we should come to terms with Wittgenstein while preserving a role for privacy. Here I am only concerned with the way in which, according to Murdoch, Wittgenstein's ideas about privacy have influenced the philosophers she identifies as representative of the Prevailing Picture (i.e. Hampshire, Hare, Ayer, Ryle, Austin), even though these ideas, in fact, leave completely open the questions that those philosophers somehow took them to settle.

Murdoch's schematic reconstruction of the Wittgensteinian line of reasoning from which her contemporaries have, in her view, drawn unwarranted conclusions about mental concepts has two prongs. The first prong draws attention to the fact that the sort of inner objects that philosophy has traditionally taken to be what constitutes inner life do not play any role in the practices involving the use of a concept (any concept, mental or non-mental) because these practices rely on what people say and do, not on any piece of inner imagery. Furthermore, and this is the second prong, the idea that we have access to these inner objects through introspection is problematic from both a phenomenological and a logical point of view. In other words, the issue is not just that the inner objects that had been thought to be the instantiations of mental concepts are useless, but that there is no evidence that they exist.

Murdoch's diagnosis of what goes wrong with the genetic argument turns on the fact that what she calls the first prong of the argument has understandably received more attention than the second. If something can be shown to be irrelevant, the question of its existence becomes less pressing. But Murdoch's contention is that "because something is no use it has to hastily concluded that something else is not there" (Murdoch 1970: 10). What Murdoch means here is that whereas Wittgenstein has persuasively shown that *certain mythical inner objects* that have been traditionally postulated have no role in our practices, his work has been taken, mistakenly, to license the conclusion that *inner life just as such* does not exist or, at any rate, cannot be anything more than a "shadow" of public acts.

According to Murdoch, the result of the influence of the genetic argument is a conception of inner life as constituted by relatively rare and elusive episodes that do not contribute to the applicability of mental concepts and whose very independent existence is in question. This conception of inner life clearly seems to require moral thought to focus on public behavior and on

whatever, if anything, can be made sense of as its byproduct. Furthermore, the observable behavior is taken to derive its meaning from a network of public criteria outside the control of the subject that belong to a “hard” and “impersonal” world, i.e. a world which does not provide any guidance for the conduct of the subject because it lacks any specific connection with her. This way of understanding the condition of the moral subject cuts her off not only from the contexts in which her life unfolds, but even from what one might (pretheoretically) have viewed as of as the materials of her own personality.

As we can see, the genetic argument leads quite directly to the existentialist ideas of a world devoid of any moral significance and of a subject identified with its capacity for choice because its self has been eroded by the re-description of its other capacities to relate to the world as belonging to a hard and impersonal mechanism.

Rather than merely providing support for the Prevailing Picture, the genetic argument presents it as something that logic and careful consideration of phenomenology force us to accept, at least in so far as we want to keep a subject in the picture at all. But, once again, what we find helpful and illuminating in Wittgenstein’s treatment of privacy has nothing to do with the very idea of inner life, only with a very specific conception of it. In criticizing the “position” characteristic of contemporary moral philosophy, Murdoch is reclaiming the possibility of considering different understandings of inner life from the one targeted by Wittgenstein.

The genetic argument is only one of the two main sources of the apparent compulsoriness of the Prevailing Picture individuated by Murdoch. The other one is the view she refers to as neo-Kantianism or Kantian liberalism.

1.2.2 Kantian Liberalism as a Misunderstanding of Freedom

As I have mentioned in passing in §1.2.1, Murdoch wants to give inner life a *central* place in our understanding of moral thought. Her criticism of the Prevailing Picture is not that, in focusing exclusively on publicly observable behavior, it fails to take into account something which is *also* part of moral life, namely, private thoughts of a certain kind. Her point is rather that the understanding of moral life as a whole depends on the question of privacy:

[U]pon the question of “what goes on inwardly” in between moments of overt “movement” depends our view of the status of choice, the meaning of freedom, and the whole problem of the relation of will to reason and intellect to desire. (Murdoch 1970: 9)

According to Murdoch, since the Prevailing Picture fails to adequately address the question about inner life, it also misrepresents the will with its operations and features because a realistic portrayal of the will depends on an adequate answer to that question. The Prevailing Picture, in other words, misrepresents not just the private dimension of moral life, which it brushes aside as irrelevant, but also the publicly observable one on which it exclusively focuses. In the absence of the context provided by an adequate account of what goes on before acts of choice, Murdoch argues, we can’t really recognize these acts as operations of a human will or as manifestations of freedom.

Among the reasons why the Prevailing Picture nonetheless came to seem attractive, there is its coherence with an independently appealing view, that is, a conception of reasons as, just as such, impersonal and public. This conception of reasons belongs to what she calls neo-Kantianism or Kantian Liberalism, a view that, like Wittgenstein’s critique of private objects, gains its authority from its usefulness in connection with certain specific problems. Murdoch concedes that Kantian liberalism has, in fact, inspired laudably antiauthoritarian political thought precisely by

locating the source of value in the will and, at the same time, denying the existence of any structure of value in the world. And yet, as she immediately adds, a good political philosophy is not necessarily a good moral philosophy (Murdoch 1970: 78-79).

One might be tempted to say that it is unsurprising that the two things should come apart for Murdoch given the importance she attributes precisely to the fact that, though moral life certainly includes our outward dealings with one another in the public sphere (i.e. the domain that largely overlaps with that of political philosophy), it is not exhausted by them. But this would be a mistaken interpretation of her concession to the merits of Kantian liberalism in political thought. The concession should not distract us from the fact that, for Murdoch, the image of the human predicament on which Kantian liberalism relies is not adequate, not even within a limited domain. The sense in which it counts as “good” is that it has been undeniably helpful in countering oppressive uses of a *particular* conception of the world as informed by values. For Murdoch, this important historical function does not change the fact that Kantian liberalism renders the idea of choice unintelligible. I will come back to this topic in §1.3.

For Kantian liberalism, as construed by Murdoch, reasons for acting are impersonal and public, because a subject’s having any such reasons amounts to having performed acts of choice that anyone else might have performed. One comes to have certain reasons for acting by attributing value to certain areas of facts, where the landscape of facts, by its own nature, is equally and unproblematically in view for any other reasonable person and, most importantly, normatively silent.

Murdoch puts this understanding of reasons in relation to a very general background conception according to which morality is incompatible with privacy because what is private is outside the reach of ordinary rational criticism. According to Murdoch, though, the idea that

something can be outside the scope of *this* kind of criticism without being outside the domain of meaningful use of concepts altogether is completely intelligible from an ordinary perspective. What makes it seem problematic is the conviction that the distinctive character that concept applications take in scientific contexts provides the model for *any* concept application. Here, again, she invites us to consider that this is not something we need to accept:

Scientific language tries to be impersonal and exact and yet accessible for purposes of teamwork; and the degree of accessibility can be decided in relation to definite practical goals. Moral language which relates to a reality infinitely more complex and various than that of science is often unavoidably idiosyncratic and inaccessible. (Murdoch 1970: 33)

Murdoch recognizes that some kinds of human action which are of interest for moral philosophy, such as paying one's bills, belong to practices which, like science, are inherently cooperative, and, therefore, that it is appropriate with respect to such actions to expect that, at least in the central cases, they be performed for ordinary and therefore easily accessible reasons. Her criticism of the conception of morality informed by Kantian liberalism is that it takes into consideration only cases of this sort, in a way that is by no means required by respect for science and is, if anything, indicative of a lack of appreciation of its distinctiveness as an area of discourse.

What the restricted focus on simple actions performed for public reasons (such as those that have to do with contracts and promises) prevents one from appreciating is the fact that the possibility and importance of converging on certain descriptions of facts is to be considered an exception. As Murdoch puts it: "We learn through attending to contexts, vocabulary develops through close attention to objects, and we can only understand others if we can to some extent share their contexts. (Often we cannot.)" (Murdoch 1970: 31) The model leaves out the sort of case in which moral reasons are not public in any straightforward sense because accessing them

requires an effort to confront the complexity and variety of reality by modifying and refining both one's concepts and one's way of applying them in response to highly specific contexts.

Kantian liberalism therefore, according to Murdoch, ignores the most demanding part of moral life. As a result, its main claims (i.e. that the will is the sole source of value, that freedom is the main virtue, and that action, choice, and responsibility are the main concepts of morality) acquire a quite modest significance despite their apparent solemnity. Within this framework, the moral subject is represented as unhindered in her gathering of the facts and equally unhindered in the production of her choices. She is "highly conscious" (i.e. her appreciation of facts is a quite straightforward and simple operation) and "self-contained" (i.e. none of the facts open to her appreciation enter into constituting her choosing self; Murdoch 1970: 34).

This representation of the moral subject is not compulsory and, furthermore, contradicts something that for Murdoch has the authority of a self-evident truth: that moral life constitutively involves tasks we can only strive toward but not accomplish or, equivalently, that there is an ineliminable role in moral life for the idea of perfection (Murdoch 1970: 27). Erasing this fact, by Murdoch's lights, results in a moral philosophy which is not only objectionably unambitious, but also unrealistic. This is why Murdoch thinks that its pervasiveness calls for an explanation, which she finds in the way in which the different doctrines operative in it both gain authority from their usefulness in contexts different from moral thought and seem to support each other's status as components of the ultimate picture of moral thought.

The idea that reasons are by definition public, i.e. specifiable in terms of facts easily accessible to anyone, which Murdoch attributes to Kantian liberalism, is a natural complement to the idea, examined in §1.2.1, that there is never anything elusive in the meaning of mental concepts, since such meaning is fully specifiable by public criteria. The two ideas, or sets of ideas,

seem to provide justification for each other in a way that makes appropriate Murdoch's use of the image, mentioned above, of the "happy and fruitful marriage."

1.2.3 Uses of Freud and the Failure to Take the Historical Individual into Account

The "happy and fruitful marriage" of Wittgensteinian logic and Kantian liberalism is, in the formula Murdoch uses to characterize the Prevailing Picture, "solemnized" by Freud. The suggestion here is that the combination of the two conceptions, respectively, of mental concepts and of reasons for acting outlined above is made stronger and more sophisticated by the contribution of psychoanalysis.

Murdoch has here in mind a specific theory of personality which she takes to be representative of the sort of philosophy informed by the Prevailing Picture and, furthermore, to provide an explicit articulation of this picture rather than tacitly assuming it: the one put forward by Stuart Hampshire.⁷ According to Murdoch, Hampshire's appeal to psychoanalysis brings into view both that the Prevailing Picture needs resources to neutralize a threat and that these resources are not available to it.

The allegedly Wittgensteinian logic and the form of Kantian liberalism that seem to support the Prevailing Picture and present it as inescapable share a very important feature, i.e. the aversion to the idea of privacy. What is private is, according to the Wittgenstein-inspired line of reasoning examined in §1.2.1, either irrelevant or non-existent and, for the kind of liberalism discussed in §1.2.2, excluded from what can count as a reason for action. These doctrines rely on impersonal

⁷ Murdoch discusses *Thought and Action* (1970) and "Disposition and Memory" (1962)

sets of rules that fix, respectively, the meaning of words and actions and what are the facts available as materials for the formulation of reasons.

The devastating problem that Murdoch sees for doctrines of this kind is the observation that what they aim to be able to account for are aspects of the life of a human being, i.e. an historical individual. For such an individual, except for very specific situations, words have personal meanings as a result of the process of constant reassessment which characteristically goes on during the life of a human being. Furthermore, and again except for very specific situations, the facts available to someone are not the same as those available to someone else. This is not just because the spatiotemporal contexts in which distinct individuals live their lives are only partially overlapping, but also because the quality and direction of their attention is different. The point might be put by saying that each individual is connected to the world in a very specific way and, therefore, the world she inhabits is *her* world: it is a world that she shares with others only in the respects in connection with which there is a point in making it happen and an opportunity to make it happen. Murdoch writes:

Freedom [...] cannot here be separated from the idea of knowledge. That *of* which it is knowledge, that “reality” that we are so naturally led to think of as what is revealed by just “attention”, can of course, given the variety of human personality and situation, only be thought as “one”, as a single object for all men, in some very remote and ideal sense. (Murdoch 1970: 37)

This talk, however, should not be taken to suggest that one makes up one’s own world, or that the world somehow bends under the pressure of the subject’s attitude. Nothing could be farther from Murdoch’s thought and, for that matter, from our experience as human beings, which Murdoch aspires to capture. A human being inhabits her world in the sense that the collection of facts she

confronts is something she conquers through her own efforts, not something that is available to her just by default.

These considerations are examples, and in fact the most central examples, of Murdoch's appeal to simple facts of human life. It is in the light of these considerations that the existentialist idea of the moral subject as a center of pure will confronting a hard and impersonal system of facts and rules starts to reveal its inadequacy. Once we remind ourselves of the fact that human beings are historical individuals, we have to take into account features of mental concepts that dislodge, at the same time, both the idea that the self has no substance and the idea that the world it confronts is hard and impersonal. Since every human being has a unique history, in fact, her conceptual repertoire is, at some level, always changing and unique to her. These two features of the conceptual repertoire of a human being, in Murdoch's view, are reflected by corresponding features both in the self (which turns out to have internal structure) and in the world confronted by the self (which turns out not to be impersonal).

Hampshire's appeal to psychoanalysis is, according to Murdoch, an attempt to block this threat by acknowledging the historical character of human personality and, at the same time, reconstructing it with the resources of the Prevailing Picture, i.e. in terms of facts just as hard and as impersonal as any other. This sort of move consists in admitting that assessing the accuracy of certain concept applications requires taking into account the history of the individual but, also, subscribing to the idea that this is a matter of registering more facts for which psychoanalysis is, at least in principle, capable of providing the "ultimate" description, that is a description whose terms enjoy some sort of primacy over other concepts.

This way of elaborating the fact that human beings are historical individuals does not, according to Murdoch, bring about any improvement in the understanding of freedom of the

Prevailing Picture. The history of the individual is conceptualized as, so to speak, a further layer of hard and impersonal facts. The inclusion of this extra layer in the Prevailing Picture, then, adds complexity and detail, but does not change the most important features of its representation of the self (which remains a mere source of choices operating from outside the world) or of the world (which remains a mechanical system lying outside the control of the individual).

Murdoch argues against this understanding of the historical character of the human individual by pointing out that nothing requires us to accept it, apart from an unwarranted assumption, namely that, if something is to count as a fact at all (i.e. part of reality), it has to have the hard and impersonal character of scientific facts. This understanding of the historical character of the human individual is not only non-compulsory, but also implausible. It implies that the only expressions of individual personality, if there are any, are interventions of the will in a mechanism that comprises everything except for the will itself, whereas in fact the efforts we make to see the world properly and their long-term results are themselves such expressions. There is no reason why we should not take the way in which someone reads a situation as any less an expression of her personality than the choices she makes in it, if a choice is called for at all.⁸

⁸ For a recent and persuasive defense of the idea that we are responsible for features of character (both moral and non-moral) that have to do with the way in which we see the world rather than with choice (such as our taste and humor) see Susan Wolf (2015: 130) “Responsibility Moral and Otherwise”

1.3 THE COGNITIVE AND THE PRACTICAL

I want to highlight two important elements that have already emerged in my overview of Murdoch's criticism of modern moral philosophy in *The Sovereignty of Good*. The first is that Murdoch's entire critical project depends on an idea that she formulates in different ways but primarily by saying that there is, in our life, a role for the idea of perfection. The second is that the Prevailing Picture is not merely partially, but overall inadequate. I want to highlight these two elements because I think that Murdoch's text is occasionally misleading with respect to them.⁹

In particular, I want to block a temptation that seems to be operative in some contemporary readers of Murdoch's work, who read Murdoch as having a special interest in some very peculiar cases belonging, so to speak, to a certain niche of moral life, namely that constituted by episodes of transformation of one's vision that do not result in either action or choice. This misunderstanding suggests that Murdoch's criticism requires us to adjust the Prevailing Picture or maybe to supplement it without really questioning its core and, particularly, its adequacy with respect to the vast areas of cases that do not share the quite distinctive features of the cases belonging to the niche.

It is important for my purposes to emphasize the radical character of Murdoch's criticism of modern moral philosophy because a proper understanding of the alternative approach that she tries, successfully by my lights, to bring back into view makes it more clearly recognizable as a

⁹ I have in mind, particularly, a passage in "The Idea of Perfection" in which Murdoch writes: "Of course choice happens at various levels of consciousness, importance, and difficulty. In a simple, easy, unimportant choice there is no reason to regard 'what goes on inside' as anything beyond the obvious sequence of reason, decision, action or just reason, action: and such choices might be properly regarded as 'impersonal'" (Murdoch 1970: 34:35)

tool in the service of the very same aims that have been pursued by relying on the Prevailing Picture.

1.3.1 Understanding Choice and Freedom

The main move that Murdoch exploits against the Prevailing Picture is a reminder, something she takes anyone to be able to come to recognize in so far as she is (or can set herself) free from philosophical prejudices. This is the claim that, though they can take many different forms, moral tasks characteristically are a matter of trying to see individuals and situations, i.e. things in the world, for what they really are, and that these attempts are very difficult and necessarily imperfect given the elusive character of reality. The determination of the will, according to Murdoch, is not achieved in the central sort of case by an act of choice performed at a specific point in time, but through the slow and long process that leads us to look at the world in the way we do when choice is called for. Moral tasks are, characteristically, a matter of striving toward perfection thought of as a completely just and discerning vision of the circumstance we face.

Murdoch is making here a special use of the notion of “seeing” in which it acquires the moral sense of being just and discerning rather than merely accurate in appreciating the features of an individual or a situation. This appreciation, for Murdoch, is not an uninteresting, matter-of-course preliminary to what is morally significant, i.e. choice, but is morally significant both in itself and because, when properly understood, it is not *followed* by choice but *culminates in* choice (if choice is involved in the morally relevant situation at hand).¹⁰ Choice, if conceived as the

¹⁰ About the case in which choice is required, Murdoch draws attention to the fact that our habitual objects of attention make any identifiable effort of the will unnecessary, since the identification of, say, intervening to stop an aggression as the thing to do has been in the making

movement of an isolated will that has surveyed a reality separate from itself and devoid of value, is not recognizable as the determination of our will. The intelligibility of this idea, i.e. that of determination of the will, requires the context provided by the work that allows us to *discover* the structures of values. This discovery is made possible by the connections between the world and the subject and *generates* new ones. The complex of these connections is what gives substance to the self: what Murdoch calls “the fabric of being.” As Murdoch writes: “I can only choose within the world I can *see*, in the moral sense of ‘see’ which implies that clear vision is a result of moral imagination and moral effort” (Murdoch 1970: 35-36). The exercise of moral imagination and moral effort, in turn, is characterized in terms of specialized uses of words, that is, uses of words that go beyond mere compliance with the public rules for their application in that specialized uses *take into account* and *continue* the process of deepening that concepts undergo as we go on with our lives.

A subject’s concepts, then, according to Murdoch reflect the foci of her attention, her way of identifying what matters and how it matters. Defending the idiosyncratic character of language as something important for moral thought is, in Murdoch’s framework, inseparable from the idea that the world we inhabit is one in which choice is possible because it is possible to come to see in it structures of value.

This understanding of choice and freedom does not exclude the possibility of acting or judging on the basis of public reasons, but it puts it in a completely different light. When we act on such reasons, it is not because we can, on the one hand, all access the facts of a hard and

during the entire time the subject has been paying attention to, say, the sadness of an indifferent society, bystander syndrome, and so on. One way to put the point is saying that the fact that stopping the aggression is the thing to do will strike the subject as just one more fact about her situation, not as an option to deliberate about. See, for instance, Murdoch 1970: 36.

impersonal world and, on the other, come to be bound by principles that have their source elsewhere and draw their authority from sources other than what individuals and situations mean for us.

There is an alternative reading of public reasons which I think follows from the general understanding of the relation between the cognitive and the practical that Murdoch presents us with. The capacity to act for public reasons, according to this reading, depends on the fact that in a certain historical moment and in virtue of having certain shared objects of attention, certain groups of people come to see (or to believe that they are seeing) the same structures of value or sufficiently similar structures of value in certain specific regions of the world they inhabit along with other people. Therefore, the part or aspect of one's way of looking at things that makes available to one a reason to act or judge can be sufficiently well-established and widespread to be called public.

One way to put the point, which has been helpfully articulated by Piergiorgio Donatelli (2012), is to say that neo-Kantianism and Utilitarianism have to a large extent misunderstood the character of their own contribution to the achievements characteristic of modern democratic societies that is (paradigmatically but not exclusively) the possibilities for choice that, in these societies, are open to an individual in areas such as the beginning and the end of life and sexuality. These two large families of theories that, as we have seen, Murdoch brings together as variants of the Prevailing Picture have represented themselves as engaged in the effort to liberate moral thought from the idea that its task is tracking the structure of values of traditional Christianity in areas like those I have mentioned and elsewhere. This structure of values was no longer able to work as an adequate background for moral thought because it started, and continued, to collide

with the ways in which individuals and situations had come to be meaningful for us as a result of different large-scale processes of transformation in our intellectual, religious, and social lives.

Neo-Kantianism and Utilitarianism have taken themselves to be leaving behind, so to speak, an empty space: the hard and impersonal world which is so central in Murdoch's discussion. The proponents of these theories have imagined that what they had been doing was bringing about and protecting the achievements of modern societies by appeal to universal principles, such as freedom and autonomy or utility, thought of as independent of *any* structure of value in the world. What these theories failed to realize, from a Murdochian perspective, is that the same rich masses of human responses to individuals and situations—the new ways of finding these things meaningful that had collided with the traditional Christian order and created the preconditions for the elaboration of the modern moral outlook—were themselves purporting to reveal the structure of value in the world, or rather fragments of such structure.

1.3.2 A Contrast Case: Anscombe's Modern Moral Philosophy

The idea of a structure of value has been implicitly or explicitly present in my discussion so far. If there is such a thing as seeing in the moral sense of “seeing”, there must be something there to be seen in this sense. If the characteristic tasks of the moral life of human beings have to do with coming to see things as they are in a way that is able to bring about the determination of their will, reality can't be thought of as hard and impersonal. Reality has to be thought of as, in fact, bearing a structure of value that can be discovered. There are, however, deeply different attitudes one might take toward the idea that one's moral thought is a claim of knowledge. I want to start to bring into view what Murdoch means in endorsing it by exploiting a contrast with a philosopher who, in the same years during which Murdoch was writing the essays collected in *The Sovereignty of Good*,

was also formulating a criticism of modern moral philosophy in equally radical tones and, to some extent, analogous terms: Elizabeth Anscombe.¹¹

In her essay “Modern Moral Philosophy,” Anscombe does not focus on the notion of choice but on the modern purportedly moral uses of the word “ought.” It is of these uses that Anscombe says what Murdoch says about choice, namely, that, contrary to what her contemporaries seemed to believe, it can only be intelligible in the presence of a structure of value. Anscombe writes:

Hume discovered the situation in which the notion ‘obligation’ survived, and the word ‘ought’ was invested with that particular force having which it is said to be used in a ‘moral’ sense, but in which the belief in divine law had long since abandoned: for it was substantially given up among Protestants at the time of the Reformation. The situation, if I am right, was the interesting one of the survival of a concept outside the framework of thought that made it a really intelligible one. (Anscombe 2005: 177)

According to Anscombe, the belief in divine law had been for centuries what provided the “framework of thought” within which moral uses of “ought” could acquire a sense. For those who subscribe to an ethics organized by divine law, the requirements of such law are identical with the requirements of the virtues, i.e. the excellences that fix what it is to be good *qua* human beings, as opposed to, e.g., *qua* logicians. In this way the divine law puts in place and guarantees, to go on with the terminology I have been using, a structure of value. There are objective facts about what it is to be good *qua* human being, and these facts make it the case, for example, that justice and courage are part of a flourishing human life and that certain individuals and certain actions are, for

¹¹ In constructing this contrast, I draw on Donatelli 2012. However, he is more critical than I am of very idea of a distinction along the lines of Anscombe’s between activities that go against human flourishing and activities that go against human life as such. I discuss in Chapter 2 some of the things one might mean by going against human life.

instance, just or courageous (and therefore good) and others unjust or cowardly (and therefore defective).

For Anscombe a divine law ethics is, in the modern world, in general unavailable. Only some very specific groups, such as Catholics like herself, can find moral uses of “ought” intelligible by relying on this kind of ethic. She also thinks, though, that the teleological understanding of human life is independent of the belief in a divine law and, though it can’t make sense of specifically *moral* uses of “ought” or of *moral* obligations, it can confer meaning on the idea of someone being good *qua* human being, i.e. an idea of human flourishing. Whereas modern science has definitively excluded teleological concepts from the understanding of the physical world at large, it is not clear that they can be similarly excluded from the understanding of the cognitive and practical capacities of human beings. A conception of what the distinctively human powers and faculties *are for* then, according to Anscombe, is available as a framework of thought able to confer a meaning on the obligations and prohibitions that we normally call “moral,” which she thinks we should rather think of as “connected with the virtues” while dispensing of the word “moral.” Since (1) I don’t find Anscombe’s motivation for dropping the word “moral” persuasive, and (2) she concedes that it is virtually impossible to do so I shall, in what follows, continue using the word as equivalent to “connected with the virtues” and not in Anscombe’s preferred sense which entails dependence on a divine law.

The teleological framework of thought that she finds in Aristotle, i.e. the idea of the human life form as setting a norm with respect to which virtues can be identified, though, is a very abstract structure, and it is far from sufficient to equip Anscombe with all she needs to bring about her program of regaining in non-religious terms the obligations and absolute prohibitions characteristic of Catholicism. In order to do so, she fleshes out this framework by appeal to a very specific set

of concepts that tradition has formed in attending to and elaborating the importance of things such as sexuality and mortality.

Anscombe's is a complicated program which I am not here interested in analyzing in detail. From the perspective of the contrast with Murdoch that I want to draw, what matters is a certain rigidity in her understanding of the status of the particular conceptual framework she takes to make the structure of value accessible for us. I'll articulate this point in the rest of this subsection in order to introduce the approach I find in Murdoch as a superior one—one which retains Anscombe's connection between the possibility of genuine moral thought, or thought about the virtues, and the commitment to the existence of a structure of value in the world, but lies, so to speak, at the opposite end of the spectrum in terms of awareness of the fallibility of our attempts to come to have it in view and of its consequences.

For Anscombe, people who do not believe in a divine law might recognize what she takes to be the genuine requirements of virtue. But this possibility depends on whether their historical circumstances allow them to inherit a very specific set of concepts that Anscombe takes to disclose what I have been calling the structure of value of the world. The alternative is the loss of some part of the human form, i.e. a life that goes against its own form. The case of Anscombe's thought about the relation between sex and marriage is instructive. She writes in "Contraception and Chastity":

[W]e do not invent marriage, as we might invent the terms of an association or club, any more than we invent human language. It is part of the creation of humanity, and if we're lucky we find it available and can enter into it. If we are very unlucky, we might live in a society that has wrecked or deformed this human thing. (2008: 185)

This passage occurs within a discussion aimed at establishing that the use of contraception within a marriage counts as deforming marriage, which is, for Anscombe, “humanly speaking, the good and the point of a sexual act” in a way that is connected with its being a reproductive act (Anscombe 2008: 185). Contraception, therefore, deprives sexual acts of the context that makes them human acts, i.e. acts characteristic of a recognizable human life.

This is a very important thing for Anscombe because it makes of what she calls “contraceptive intercourse” a very special kind of failure to comply with the requirements of virtue. Contraceptive intercourse is the object of an absolute prohibition because it has to do with a virtue, chastity, whose requirements articulate the human form of life itself rather than, as in the case of e.g. the virtue of sobriety, indicating what counts as flourishing for an individual bearing it. Operative in Anscombe’s thought is a distinction between these two kinds of virtues, which she calls respectively *mystic* and *utilitarian* (Anscombe 2008: 187). What I referred to, earlier in this subsection, as the rigidity of the structure of value Anscombe invokes to make sense of human flourishing has to do neither with the fact that she is deeply confident in such structure, nor with the distinction I have just mentioned. These are things that might be part of *an* understanding of human life. Neither am I accusing Anscombe of invoking an external foundation for her views in moral philosophy. She is openly committed to defending them in a framework, the neo-Aristotelian one, which is independent of the belief in a divine law. What goes wrong with Anscombe is more something along the lines of the sort of attitude that John McDowell considers as a likely explanation of Aristotle’s taking for granted his own moral outlook: an insufficient sensitivity to the idea that “ethical thinking, just like any other thinking, is under a standing obligation to reflect about and criticize the standards by which, at any time, it takes itself to be governed” (1996: 81). McDowell calls it “smugness.”

The neo-Aristotelian teleological understanding of human life, in Anscombe, goes together with a very specific set of concepts which she takes to be uniquely capable of adequately articulating the structure of value of the world, thereby making it possible to speak of virtues and of actions as right or wrong in the ways that the concepts of the virtues identify.¹² But this is not the only way to develop the idea that moral thought stands in need of a structure of value in the world. What Murdoch helps us to bring back into view is a possibility for practical thought that both what I have called the Prevailing Picture and Anscombe, for different reasons, overlook.

The loss, or rather the active demolition, of the premodern understanding of our place in the world does not leave us with the task Anscombe indicates for moral philosophy. Moral thought, and moral philosophy as part of it, is not in the business of retrieving from its remnants in the tradition some adjusted version of the meanings that that understanding had placed on the facts of human life, as if this constituted our only chance (if it is such chance at all) to live a good human life. But neither does this historically important transition leave us in a world without any structure of value, facing the task of injecting, so to speak, meaning into the world we inhabit from some separate sphere located outside that world.

We can, for instance, continue to see the fact that human beings reproduce sexually as a salient and meaningful fact about ourselves and yet refuse to see the use of contraception within marriage as a way of “wrecking” or “deforming” it. We can indeed, still within a perspective that

¹² See, in this connection, Anscombe (2005): 71. There, Anscombe writes: “our sexual activity and reproduction is all tied up with our intellect, our not merely animal emotions and our aesthetic feelings. Reason and love enter into most and certainly into all characteristically human exercise of this vital function. Hence marriage and the celebration of and awe before procreation and pregnancy.” Up to a point, Anscombe makes claims that are just characteristic of neo-Aristotelianism or second nature naturalism, but the word “hence” signals what I have called “rigidity”: a scarce sensitivity to the obligation to criticize the tools purporting to disclose the structure of value of the world.

acknowledges the meaningfulness of sex within human life, refuse to think of marriage as a “human thing” in a sense that implies that other ways of engaging in sex are not properly human or contrary to human nature. We can, in other words, retain Anscombe’s Aristotelian thought that an action is good in so far as it belongs to an activity which is properly human in that it constitutes the exercise of a capacity that contributes to human flourishing, without endorsing the view that any particular set of conceptual resources present in the tradition has already fixed once and for all what human flourishing amounts to (and, with this, what are the properly human activities and which traits of character make one perform them in an excellent way).

Anscombe sees no alternative to an understanding of human life that tracks very closely that of Catholicism other than plain and simple loss of our humanity or, at least, of the possibility of its flourishing. However, this is in no way essential to the approach to ethics centered on the notion of virtue which she has contributed to introduce into the contemporary debate. Michael Thompson, for instance, in “Apprehending Human Form,” openly acknowledges that—although our claims to the effect that something is, for instance, unjust commit us to a certain stance on how things are, in a certain respect, with human form—it is only reasonable to take into account that our grip on such structure is only as good as our upbringing and reflection. He writes:

In representing my propositions as possessing so-called normative authority [...] I represent them not simply as manifestations of the form I bear, as all my thoughts (and heartbeats) are, but as characteristic of the form I bear. I thus represent myself as in this respect in possession of a sound practical understanding *qua* bearer of this form. [...] Of course we have no way of judging what practical thoughts and what range of upbringings might be characteristic of the human, and sound in a human, except through application of our most fundamental practical judgments—judgments about what makes sense and what might count as a reason and so forth. And these are judgments each of us must recognize to be the result of his own upbringing and reflection. (Thompson 2004: 73)

Anscombe seems not particularly sensitive to the possibility that the concepts she takes to bring into view human form and human flourishing might be inadequate. What I find in Murdoch is a version of the approach to moral thought centered on the notion of virtue that is largely organized precisely by sensitivity to this possibility. Murdoch's view, while having important elements of commonality with Anscombe's, is interesting and distinctive precisely in this respect. In Murdoch, the acute sense of the elusiveness of the reality which is the source of the requirements of virtue and of the fallibility of our attempts to see reality it is at the center of the scene. Moral thought (and moral philosophy as part of it) is, in her view, primarily concerned with investigating the sources of our failures at seeing reality for what it is and the resources for doing better. This, I think, enables us to understand certain very prominent and distinctive elements in Murdoch's thought, such as the importance attributed to the idiosyncrasy of our moral uses of words and to the historical character of human individuals. Both these elements might strike one as in tension with her frequent declarations of realism in moral thought. The following chapters are devoted to bringing into view how they are not just tolerable in a realist framework but actually required by it.

2.0 THE CONCEPT OF HUMAN BEING IN MORAL THOUGHT

The general framework that Murdoch and Anscombe have contributed to bringing back as a possibility for ethics is one according to which some practical questions are not answerable in terms of considerations of compliance with conventions or efficacy with respect to the desires one happens to have. One might, for instance, wonder whether she should attend a funeral, or forgive an offense, or report a case of misconduct when there are no doubts about what would best satisfy her desires or what the conventions require of her. The point can be put naturally by saying that one sometimes wonders what to do, or assesses someone else's doings, *as a human being*, and finds an answer in a conception of what it is to do well for such a being which stands a chance of being true. This perspective is what, according to the framework in question, makes practical questions moral.

In the context of philosophical reflection on morality, though, the appeal to the notion of humanity is often looked at as misleading or plainly out of place. We have, then, an instance of a phenomenon which Murdoch is interested in and which constitutes the main topic of this dissertation. This phenomenon, as I have explained in Chapter 1, has the following structure. There is something that, independently of philosophy, we take to be an important aspect or dimension of moral thought because our own way of thinking about how to live has, in fact, that aspect or dimension, or because thinking morally in that way constitutes a possibility for us (maybe one that we see actualized in other people). But, when we consider moral thought from a philosophical point of view, the same aspect or dimension becomes irrelevant or even invisible. This is not

because of any inherent limitation of philosophy, but because of a tendency to take some of its theories to be “ultimate” in the sense illustrated in Chapter 1.

As I have argued in Chapter 1, Murdoch’s criticism of the moral philosophy of her times is to a large extent aimed at regaining the possibility of comparing genuinely different ways of understanding the place of the moral subject in the world or, as she puts it, “producing a rival soul-picture” for the one she saw as occupying, with its variants, the entire space of visible alternatives. In this picture, the world is represented as a hard and impersonal mechanism inhospitable both to values and to subjectivity. As we have seen, an important part of Murdoch’s line of reasoning in *The Sovereignty of Good* is her objection to the idea that this picture is made compulsory by science and logic. Once one comes to see that the picture is not compulsory, two new elements enter the scene: a soul-picture that competes with the existentialist one and a criterion for adjudicating this competition. The soul-picture is one in which the moral subject is not a center of pure will intervening from outside in a world devoid of value but the historical human individual inhabiting her world, and the criterion has to do with the relation that each of the alternative pictures has with a certain kind of knowledge: a kind of self-understanding human beings are able to achieve.

In this connection, Murdoch writes, at the end of “The Idea of Perfection”:

Philosophers have always been trying to picture the human soul, and since morality needs such pictures and as science is, as I have argued, in no position to coerce morality, there seems to be no reason why philosophers should not go on trying to fill in a systematic explanatory background to our ordinary moral life. (Murdoch 1970: 43)

A couple sentences later she adds:

[T]he sketch which I have offered, a footnote in a great and familiar philosophical tradition, must be judged by its power to connect, to

illuminate, to explain, and to make new and fruitful places for reflection. (Murdoch 1970: 44)

In this chapter I will take some steps toward the characterization of the kind of knowledge presupposed by the many cognitive concepts in these two passages and of the relation with such knowledge that counts in favor of a picture of the soul. I'll do so by focusing on a class of forms of interpersonal wrongdoing which includes both those characteristic of localized explosions of violence, such as genocides, and those proper to different kinds of systemic oppression, such as sexism and racism. I think that in connection with these phenomena the way in which a conception of human life works as the background of moral thought is especially transparent.

2.1 A CRITICISM OF THE MORAL USE OF THE CONCEPT OF A HUMAN BEING

2.1.1 A Familiar Concept Under Scrutiny

In thought and speech about the most appalling forms of interpersonal conduct, the reference to the notion of humanity is typically made explicit by the use of a cluster of concepts all standing in some kind of opposition to it. It is, in fact, very natural to characterize genocidal threats and violence, as well as racist or sexist ones, as *dehumanizing*. The thrust of the criticism I am going to consider is the claim that describing these forms of wrongdoing in these terms either is not explanatory at all or is an explanation flawed by a major blind spot.

I argue that this kind of criticism, as developed e.g. by Kate Manne (2016), and certain strategies for replying to it, e.g. the one pursued by David Livingstone Smith (2016), presuppose an understanding of what the notion of *a human being* can possibly mean which is organized by assumptions that Murdoch worked to dislodge, but that are still widespread in contemporary philosophy. If our understanding of the concept of dehumanization were in fact constrained in the way these commentators take it to be, its use would indeed amount to producing a bad explanation of the forms of conduct to which it purports to apply. But if one follows Murdoch in rejecting these perceived constraints, a much more familiar and plausible interpretation of the use of this concept becomes available.

In what follows I criticize the way in which the range of acceptable interpretations of the concept of a human being assumed by those who reject its moral relevance, and I offer an alternative interpretation that they exclude as a result of certain general philosophical commitments. This alternative interpretation consists in taking the concept of a human being to signify neither a member of a biological species nor a being with this or that feature or set of features which is, somehow, inherently normative, but a bearer of the distinctive kind of life that we, i.e. we who use that concept, also bear and, as Diamond puts it, imaginatively respond to (Diamond 1991: esp. 46-47).

The activity of making imaginative sense of human life is, according to Diamond, a characteristically human activity, and it is possible for moral philosophy to be continuous with it. She shares with Murdoch the idea that respect for science does not exclude that activity as contrary to our best understanding of nature, and credits Murdoch with having drawn attention to its legitimacy and value as a method for moral philosophy and for philosophy in general (Diamond 2011). Applying this philosophical method amounts to doing what we do when we intelligently

engage with each other, i.e., treating our conception of humanity as the background that confers their specific significance on our actions and attitudes. This is an activity which is inexhaustibly complex and, for this reason, very hard to describe, but also an extremely familiar one. The main claim of this chapter is that a philosophical perspective that makes it invisible is, on that account, deeply problematic.

It's very common to use the term *dehumanizing* in connection with the kinds of wrongdoing characteristic both of episodes of explosive violence, such as genocides, and of systemic forms of oppression, such as racism and sexism. Implicit in this kind of talk is the idea of a tension of some sort between, on the one hand, wronging someone in these ways and, on the other, recognizing her as a human being. This idea, in turn, suggests a familiar explanation of the moral psychology of the kind of conduct at issue: its perpetrators do what they do because they don't see their victims as human beings. In the two contexts I have just mentioned as those in which dehumanizing conduct most typically occurs, individuals are not targeted *qua* individuals, but *qua* members of certain groups, and a very natural way of understanding how membership in these groups elicits purportedly dehumanizing conduct by certain agents is to understand this membership to be what undermines someone's status as a human being in these agents' eyes.

Some authors have taken issue with the way in which this language is generally used because it is most characteristically deployed in connection with forms of conduct that, in their view, rather than expressing or betraying a lack of awareness of the humanity of those on the receiving end, crucially presuppose it.

These authors start with a conception of how the idea of *seeing someone as a human being* must be understood, and let this conception guide them in the interpretation and assessment of the claim that certain forms of conduct stem from not recognizing certain human beings as such. Their

conclusion is that this kind of explanation is inadequate for most of the phenomena for which it is used.

I argue that we need to follow the opposite route. The role that the kind of explanation just outlined has in moral discourse should be treated as a source of guidance in our effort to understand how the concept of a human being works in the practical sphere and as a ground for criticizing the assumptions that are taken as starting points by those who want to impugn that role.

Here is an example of the sort of criticism I want to address. Adam Gopnik, writing about the historical phase known as the Reign of Terror, writes:

It is often said that terror of this kind is possible only when one has first “dehumanized” some group of people—aristocrats, Jews, the bourgeoisie. In fact, what motivated the spectacle was exactly the knowledge that the victims were people, and capable of feeling pain and fear as people do. We don’t humiliate vermin, or put them through show trials, or make them watch their fellow-vermin die first. The myth of mechanical murder is almost always only that. (Gopnik 2006)

A further example of this approach is provided by Kwame Anthony Appiah. He writes in *Experiments in Ethics* (2008: 144) that the familiar answer to the question about how outgroup conflicts degenerate into genocidal episodes is: “By persuading us that members of some other group aren’t really human at all.” Appiah claims that this view is “not quite right,” and explains:

The persecutors may liken the objects of their enmity to cockroaches or germs, but they acknowledge their victims’ humanity in the very act of humiliating, stigmatizing, reviling, and torturing them. Such treatment—and the voluble justifications the persecutors invariably offer for such treatment—is reserved for creatures we recognize to have intentions and desires and projects. (Appiah 2008: 144)¹³

¹³ Even though some of the things Appiah mentions can be inflicted on non-human animals, the forms and the point of doing them to human beings are different. A cow can be tortured but

Both passages assume that explaining a kind of conduct in terms of dehumanization entails that its perpetrators engage in that kind of conduct because they see their victims as other than human beings, e.g. as vermin. They also draw attention to the fact that many of the forms of conduct to which this explanation is supposed to apply are only intelligible as directed at beings that have distinctively human capacities. Both Gopnik and Appiah take this consideration to be sufficient to conclude that describing these kinds of action as involving dehumanization, and explaining them accordingly, is simply a mistake.¹⁴

One might be surprised at the hastiness of this conclusion, especially given that, on the one hand, the consideration that here plays the crucial role is hardly a very sophisticated or easily overlooked one, and, on the other, the contested use of the idea of dehumanization is characteristic not just of everyday speech, but also of many intellectually rigorous attempts to make sense of the most traumatic human experiences.¹⁵ I am not here trying to suggest that we should accept the

not feel the shame of having betrayed her deepest commitments in order to make it stop; dogs of certain breeds are stigmatized but they don't have to deal with the awareness of having passed their kind and its afflictions to their offspring. Also, I don't think Appiah means to deny that beings other than humans have desires. The spirit of his remark is, however, clear.

¹⁴ The passages by Appiah and Gopnik that I have just quoted are both discussed in Smith (2016). Smith defends the central explanatory claim I defend here from the same sort of criticism but does so through a different strategy. Specifically, Smith shares with the perspective represented by the two passages just quoted that the point of the explanation of inhumane conduct they criticize can only rest on some sort of empirical generalization about the causal effects of attributing certain features to an individual (see especially pp. 440-442). With this I disagree. Despite this fundamental difference, I find that Smith's work on the phenomenology of inhumane conduct is an excellent example of the sort of patient, rigorous, and nuanced observation required by that field of phenomena.

¹⁵ A paradigmatic example of this kind is Primo Levi's use of the framework of failing to see the humanity of certain individuals (and thereby destroying it) in connection with the horror of Nazi camps in *If This Is A Man* (1959). As is the case for many characteristic uses of this framework, this is not a disengaged speculative explanatory hypothesis, but the attempt, as Levi says in an interview (1986), "to bear witness to things endured and seen." Another example of

applicability of this kind of explanation just on the ground of the number or authority of those who use it. My point is simply that the reaction of Gopnik and Appiah shows a striking lack of charity. Having observed that describing the victims of show trials and humiliation as not seen as human beings, in the sense they ascribe to this claim, leads to a story about human motivation that is blatantly implausible for most of the forms of conduct to which it is meant to apply, neither of them explores, or even mentions, the possibility of interpreting that claim in some other way.

I think that this happens because of certain perceived restrictions on the kind of explanation in which the notion of a human being can intelligibly appear. These restrictions emerge in a particularly clear way in a recent article by Kate Manne, “Humanism: A Critique” (2016). Though Manne interprets the explanation that invokes the idea of dehumanization in the same way Appiah and Gopnik do, she articulates explicitly how one arrives at that interpretation. I will exploit her critical argument to identify the sources of these perceived restrictions in two assumptions: the Humean Theory of Motivation and a narrow understanding of explanation in moral psychology.

2.1.2 **Manne’s critique of Humanism**

The criticism exemplified in Appiah and Gopnik, as we have seen, is directed not to the sort of explanation that invokes the idea of dehumanization *per se*, but to a certain widely shared conception of its scope of application: one that, by their lights, takes it to encompass too much. According to this conception, this sort of explanation is generally appropriate for the entire field

authoritative use of this concept with an emphasis on the power of words comes from Maya Angelou: “The n-word was created to divest people of their humanity,” she said in a 2006 meeting with David Chapelle.

of phenomena whose paradigmatic forms are genocidal and racist threats and violence, not just for a specific region within it. The framework underlying this conception is what Manne takes as her target and refers to as “humanism” in her article. She takes this framework to pervade discourse concerning the phenomena just mentioned, and to have been explicitly endorsed, in one version or another, by many thinkers otherwise very different from one another. In the rest of the chapter, I will follow Manne in referring to the framework under discussion as “humanism,” and I will refer to the form of explanation in terms of which she specifies it, i.e. the one invoking the notion of dehumanization, as “the familiar explanation of inhumane conduct.”¹⁶

Manne’s discussion of humanism is not meant to reconstruct any particular philosophical articulation of it, but to individuate some key commitments that are typically accepted by its advocates and to combine them in a “natural, attractive package” (Manne 2016: 395). Manne’s adoption of this strategy is very helpful for my purposes because her elaboration of what it would be for a philosophical articulation of humanism to be “natural” or “attractive” brings to the fore the philosophical assumptions, about the role of the concept of a human being in moral psychology, that I want to address.

Manne formulates humanism as the view in moral psychology according to which

¹⁶ As Manne acknowledges, the term “humanism” has been used in many different ways and using it to designate the framework she criticizes is potentially confusing (2016: 390n4). This is, in fact, a reasonable concern, but my main worry in using it is not the possibility that one “ism” (i.e. a kind of philosophical theory) be confused with another. The risk I want to avert is rather that what is, primarily, a characteristic human activity, i.e. making imaginative sense of human life, might be identified with something which is primarily a kind of philosophical theory. As I am going to argue, the practical significance of the concept of a human being is something that practical philosophy has to be able to accommodate (or to convince us to reject, as the case might be). But such significance does not need to be established by argument as if it were not already a (particularly salient) part of the conceptual resources we live our lives with.

interpersonal conduct of the kind that is naturally described as *inhumane*...often stems from people's failure to recognize some of their fellows *as* fellow human beings. (Manne 2016, 390)

A crucial word in this formulation is “often.” Manne is interested neither in maintaining that the view she opposes entails that the explanation in terms of dehumanization applies without exceptions to all forms of interpersonal conduct of the kind naturally described as inhumane, nor in specifying which exceptions such a view might consistently allow for. Her main point is that humanism treats this explanation as applicable to the most paradigmatic varieties of this kind of conduct, such as genocidal and racist threats and violence, whereas, for her this is clearly wrong.

Manne thinks, with Appiah and Gopnik, that, if a form of interpersonal conduct is to be explained by appeal to the idea that the perpetrators do not see their victims as human beings, it must be at least intelligible for that conduct to be adopted toward pests or monsters or whatever sort of beings the victims are, allegedly, seen as instead. The question of the intelligibility of performing an action on an individual other than human, furthermore, is to be assessed by considering not just the physical movements that that action involves, but also its motivations and objectives. Take, for instance, subjecting dissidents to show trials. It is certainly possible to bring non-human animals in front of a tribunal and utter certain sentences before killing them. But this does not amount to subjecting them to a show trial: there simply is no such thing as subjecting non-human animals to a show trial, because there is no such thing as subjecting them to a trial.¹⁷

¹⁷ A peculiar civil ceremony of this kind was instituted in Venice in the twelfth century as part of an agreement between the patriarch of Aquileia Ulrich II of Treffen and Doge Vitale II Michiel. In exchange for the release of Ulrich himself and the twelve canons of his cathedral who had been captured by the doge, “Ulrich promised that henceforth, every Fat Tuesday, he and his successors would send to the people of Venice a bull and twelve pigs, to represent the patriarch of Aquileia and his canons. Thus began another of the most enduring civic ceremonies in Venetian history. Each year for centuries these animals were solemnly brought to the Piazza San Marco,

Analogous considerations show that the kind of physical torture perpetrated against political adversaries cannot intelligibly be inflicted on non-human creatures. The specificity of this form of wrongdoing, arguably, is not exhausted by the physical violence it involves, but is partially constituted by the objective of humiliation—an objective which does not make sense in connection with, e.g., rats.

This is what, according to Manne, the advocates of humanism must have failed to notice. Humanism, in fact, involves essentially two connected claims. The first is that we are capable of seeing or recognizing other human beings as fellow human beings. The second is that this recognition is not just a precondition of humane conduct, but also something that tends to strongly motivate us to refrain from forms of conduct naturally characterized as inhumane. The explanatory claim that Manne disputes is a natural complement to these claims. If one thinks that recognizing other human beings as fellow human beings tends to strongly motivate us in this way, she will also think that when one engages in forms of conduct naturally characterized as inhumane, it normally depends on the lack of such recognition.

Manne spells out the idea of seeing someone as a human being that is operative in what she takes to be the most charitable reconstruction of the ordinary view as a cognitive state consisting in the attribution of a set of features. She leaves the set of features somewhat unspecified, since different advocates of this view will construe the attitude of recognizing someone as human differently. However, she gives the flavor of the sort of features that she takes to be likely to appear in that set by mentioning things like having the capacity for social relations such as friendship and marriage, and the capacity to be the object of deep emotional attachments

where, under the watchful eye of the doge and his council, they were tried, convicted, and condemned to death” (Madden 2012: 106-7).

(or merely having, or having had, the potential to develop such capacities; see Manne 2016: 396). Manne, in other words, takes as representative of the features characteristically included by the advocates of humanism in the articulation of what it means to see someone as a human being capacities whose exercises, at least superficially, seem to be of a sort likely to elicit a friendly stance.

This feature of Manne's reconstruction of humanism is the central element of her discussion. It is what she thinks necessary in order to turn the thought implicit in our ordinary way of talking about conduct naturally described as inhumane into a story that is worth considering about the moral psychology of such conduct. At the same time, it is the element that makes this story vulnerable to a fatal objection.

As a subscriber to the Humean Theory of Motivation, Manne thinks that being in a world-guided state, such as having the belief that someone is a human being, does not by itself motivate one to act, or to refrain from acting, in any particular way. According to Manne, if the humanist explanation of the conduct naturally described as inhumane is to have an even *prima facie* plausibility, according to Manne, humanism needs to be reconstructed in a way that makes it capable of bridging what she sees as a gap between recognizing certain individuals as human beings and the disposition to refrain from treating them in ways naturally described as inhumane. Only once this is accomplished can the lack of this recognition be credited with a role in the explanation of conduct naturally described as inhumane. She identifies as the most promising way of bridging this gap the following line of thought:

in view of our recognition of someone's similarity to ourselves, we will be able and inclined to identify with her, or (somewhat more modestly) to take her perspective. We will subsequently often feel what we imagine she feels, or at least experience congruent or "helper" emotions (pity being one such). This being the case, we

will tend to want to be kind, rather than cruel, to her—or even to help, not to hurt, her more generally. (Manne 2016: 397-8)

The idea here is that, even though any view that takes the way one sees an individual to be sufficient to explain the way she engages with that individual is inherently defective, the view that there are characteristic causal connections between representations and behavioral responses is not. One only needs to explain what mechanism accounts for such causal connections.

If recognition of another individual's humanity can be causally connected with a feeling of empathy and this, in turn, can be causally connected with a motivation to refrain from aggression and even to offer cooperation, then the lack of such recognition can be seen as contributing to a plausible explanation of conduct naturally described as inhumane. It will be relevant to the etiology of episodes of such conduct that a mechanism that normally would have inhibited them was not operative. Here we see what it means, for Manne, to make of the humanist commitments a “natural, attractive package.” Without the appeal to empathy or a similar mechanism, she thinks, the familiar explanation of so-called inhumane conduct would be structurally inadequate as an explanation of *any* kind of conduct. When empathy is invoked, this explanation becomes, by Manne's Humean lights, a well-formed explanation, but also one that—on reflection—turns out to be false for most forms of so-called inhumane conduct.

Manne does not reject the idea that the recognition of someone as a human being *can*, in principle, cause an inclination not to hurt or even an inclination to help. Her objection is that this is very far from being the typical way in which such recognition works, because we very rarely see someone *just* as a human being. This representation is virtually always embedded in more complex representations, and while some of them tend to elicit empathy and the associated

motivations to act, others tend to cause hostility and the associated motivations to act. Manne writes:

a fellow human being is not just an intelligible spouse, parent, child, sibling, friend, colleague, and so on, in relation to you and yours. She is also an intelligible rival, enemy, usurper, insubordinate, betrayer, and so on. (Manne 2016: 399)

This is the blind spot that Manne attributes to humanism. Ascribing distinctively human features to other individuals amounts to taking them to be similar to ourselves, but, though this might present them to our eyes as “relatable,” it might also (or instead) present them as “dangerous” and “threatening” in distinctively human ways (cf. Manne 2016: 400). When one sees fellow human beings as dangerous and threatening, one will tend to be antagonistic, not friendly; and so one will be motivated not to help them and, possibly, even to hurt them. Manne thinks that advocates of humanism must, somehow, fail to notice this fact because of the motivational significance they, according to her construal of their position, assign to the recognition of someone else’s humanity.

The antagonistic dispositions elicited by seeing someone as an enemy or a rival, Manne argues, essentially involve the representation of that individual as a human being, and such negative dispositions arguably can, not just outweigh whatever empathy might have been elicited by this representation on its own account, but also prevent it from arising in the first place.¹⁸ Each of these possibilities, according to Manne, corresponds to a kind of interpersonal stance that the advocate of humanism is committed to treating as an exception, because it involves seeing

¹⁸ Manne talks about the possibility that the negative dispositions associated with representations such as *enemy* and *rival* “cancel” the positive disposition that, as she concedes, tends to be elicited by registering someone’s humanity. Since she takes this possibility to be something different from the mere outweighing of this positive disposition by a competing negative one, it is fair to suppose that “cancel” stands here for “prevent from arising.”

someone both as a human being and as a target of one's hostility. But these interpersonal stances, Manne goes on, are not exceptions. Representations of others as *rivals*, *enemies*, and the like are just as likely to shape our way of seeing individuals we recognize to be human beings as representations of such individuals as *friends*, *allies*, and the like are.¹⁹

Manne's conclusion is that conduct naturally described as inhumane is often to be accounted for by appeal to representations of the individuals one confronts like *enemy*, *rival*, and a host of other ones that are similar to them in that they presuppose humanity and are, nonetheless, hostility-inducing. This conclusion is the core idea of Manne's proposed alternative to humanism, which she calls "the socially situated view." The main explanatory claim of this alternative approach can be formulated as follows:

Socially situated view: So-called inhumane conduct often stems from representing others in ways that, at the same time, presuppose their humanity and elicit antagonistic stances. These ways of representing other individuals are exemplified by seeing them as *enemies*, *rivals*, *usurpers*, and the like. (Cf. Manne 2016: 402-3)

As Manne implicitly acknowledges, this cannot be the whole story about the conduct she wants to explain. She writes that the antagonistic stances elicited by the additional representations that she takes to be overlooked by the advocates of humanism do not necessarily lead to conduct we would call inhumane (Manne 2016: 402-3). Hostility toward other human beings recognized as such, certainly, can be both justified and properly acted upon. Furthermore, it might lead to

¹⁹ This observation is part of what Peter Strawson calls "the central commonplace" in his discussion of reactive attitudes (Strawson 1962: 190-1).

conduct that, though problematic, is not naturally described as *inhumane*, but rather as, e.g., *disproportionate*, *distasteful*, or *petty*. For these two reasons, the socially situated approach tells us at best a part of the explanatory story that Manne is after.

Despite this fact, Manne takes the main claim of her situated approach to be able of establishing her negative point, namely, that humanism is wrong about the scope of the familiar explanation of conduct naturally described as inhumane. If the most paradigmatic forms of such conduct are a specific way of reacting to other individuals *qua rivals*, *enemies*, and the like, and if these representations presuppose humanity, then conduct naturally described as inhumane cannot be traced to a dehumanizing attitude.

This line of reasoning depends on a misunderstanding of the connected concepts of recognition of someone as a human being and of dehumanization invoked by the sort of framework Manne calls humanism. These concepts do not, as she assumes we are somehow forced to think, correspond respectively to registering and overlooking features that tend to elicit cooperative behavior or, at least, to inhibit antagonistic behavior. To explain how I think we should understand them, I will draw attention, in §2.2.1, to the fact that a tendency to cause antagonistic behavior directed at a certain individual is not only insufficient but also unnecessary to qualify someone's attitude toward that individual as dehumanizing. Before doing that, I am going to argue, in §2.1.3, that the misunderstanding of the connected concepts of recognition of someone as a human being and of dehumanization that I am attributing to Manne (and the other critics of humanism I have mentioned above) is also characteristic of certain lines of defense of the humanist framework.

2.1.3 A Line of Defense for Humanism

David Livingstone Smith (2016) has produced an interesting and I think, within its scope, successful defense of what he calls “dehumanization theory,” which can definitely be recognized as a version of what I have been calling “humanism,” from the sort of critique Manne and others have leveled against it. As I am going to explain, though, there are reasons for wanting a defense of humanism to go further.

Smith’s understanding of the problem posed by forms of conduct naturally described as inhumane is that empirical psychology and various sorts of ordinary experiences tell us that it is psychologically hard to mistreat other human beings, and, therefore, when this happens in a deliberate, studied, and even apparently pointless way, this calls for an explanation. Smith acknowledges the obvious fact that human beings can obtain all sorts of (what they perceive as) benefits by killing or exploiting others, but urges, *contra* Manne, that the prospect of an advantage does not by itself eliminate or easily overcome the psychological resistance to killing or harming them. Dehumanization, as Smith understands it, is one of the solutions to what he calls *the problem of ambivalence*. Where obtaining certain perceived benefits requires killing or harming other human beings, their humanity gets, so to speak, psychologically in the way, and dehumanization is one way of getting rid of this obstacle. Manne’s observation that often what triggers the allegedly dehumanizing conduct are precisely manifestations of humanity recognized as such by the perpetrators, therefore, does not impugn Smith’s explanatory hypothesis. This hypothesis does not require that the perpetrators fail to recognize their victims’ humanity altogether, it requires only that they experience it as a psychological obstacle to overcome. Dehumanization figures in Smith’s explanation as what accomplishes this task.

This line of reasoning about the function of dehumanization, which Smith supports by reference to empirical studies, is often found in reports of direct experience of atrocities. Yang Zhanqing, a public-interest campaigner who was detained for thirty days in one of the Chinese extra-legal facilities sometimes referred to as “black jails,” offers a very similar analysis in a piece he wrote for the *New York Review of Books*. Having described the detention center’s policies of not allowing inmates to have either spoons or toilet paper, he writes:

Are these gratuitous humiliations really “useless”? Hardly. The goal is to break a person’s will, because a person of broken will is easier to manipulate, and our Center is therefore easier to run. And there is an added benefit: the wardens are, after all, human beings, born with consciences, and it can be psychologically stressful for them to have to “twist” other human beings. But if the others are less than human beings? Only semi-human? It is easier. (Zhanqing 2016)

It seems that an analysis like Smith’s can quite easily accommodate Manne’s observation that the *motivations* of the perpetrators of atrocities presuppose that they are aware of their victims’ humanity. The main problem for this kind of analysis comes from the consideration I have articulated above about the *character of the acts* performed by the perpetrators of, e.g., torture and genocidal violence. These forms of conduct, as Manne, Appiah, and Gopnik point out, are intelligible for what they are only if we keep firmly in the picture the perpetrators’ awareness of their victims’ humanity *as they engage in them*, and not only *as they are in the process of overcoming a psychological difficulty to engage in them*.

Smith’s reply is very rich and nuanced and perhaps my summary is not going to do full justice to it. However, its main idea is that not only do human beings find it psychologically hard to harm or kill other human beings; they also find it hard to conceptualize them as other than human. As *hypersocial* animals, Smith argues, we are “exquisitely sensitive to cues indicating that

others are members of our own species” (2016: 435).²⁰ Therefore dehumanization, which he takes to be precisely this conceptualization of a human being as other than human beings, does not come easily to us. When political propaganda or other mechanisms through which dehumanization is pursued push people to see members of a certain group as subhuman (e.g. rats or cockroaches), this induces in them a cognitive dissonance between what they have been brought (or have brought themselves) to believe and what they seem to be confronting. The members of the dehumanized group, therefore, come to be seen as *both human and subhuman*, rather than being straightforwardly conceptualized as belonging to a kind different from the human species.

This unintended consequence of the dehumanizing process, according to Smith, explains the distinctive kind of attitudes and behaviors toward dehumanized groups: they are not those that people are expected to have toward rats, cockroaches, and other repulsive animals, but toward *rats or cockroaches in human form*. This stance becomes intelligible, Smith argues, when we understand that dehumanization places human individuals in the realm of the metaphysically ambiguous: the domain of the impure, the uncanny, and the monstrous.

An important feature of Smith’s defense is that it does not require that the dehumanizers explicitly attribute a subhuman essence to the members of dehumanized groups, nor does it require that they explicitly deny that their victims are human. Smith explains the core commitment of his theory by saying that:

[it] requires only that dehumanizers regard those whom they torment as both human and subhuman, which might occur tacitly, explicitly, or by means of a combination of the two. Explicit assertions that one or another group of people is less than human are neither necessary nor sufficient for dehumanization. (Smith 2016: 435-6)

²⁰ Smith borrows the idea of human beings as hypersocial from Blaffer Hrdy 2009.

In this way the theory accommodates the recognition of humanity as a crucial element of its explanation, while preserving a role for the notion of dehumanization that is not merely rhetorical. Smith's claim is, ultimately, that the attribution of distinctively human capacities to the victims of atrocities does not undermine the claim that they are dehumanized by the perpetrators of such atrocities, because dehumanization amounts not to withholding the attribution of such capacities, but to *attributing such capacities in combination with a non-human essence*.

Manne has resisted this part of Smith's defense by relying on one of the main examples she uses to make her case that conduct naturally characterized as inhumane presupposes the perpetrators' awareness of their victims' humanity. She argues that, if the conceptualization of the female members of an ostensibly dehumanized group as non-human animals poses a problem for the explanation of mass rape, so does a conceptualization of them as uncanny subhuman creatures. According to Manne, we should expect these creatures to be as sexually repugnant as, or even more sexually repugnant than, non-human animals.

I think that Manne dismisses Smith's way of addressing this issue too hastily. Smith explains how the targets of dehumanizing discourse get stuck in between metaphysical realms by appeal to the quite convincing claim that other people's humanity is relentlessly present to us. The fact that attitudes that we take to be directed (at least paradigmatically) only toward human beings (such as sexual instincts) tend to survive the dehumanizing process makes perfect sense in the context of Smith's account. It brings out a lack of coherence not in Smith's construal of the dehumanizers' attitude, but in that attitude itself. The incoherence that he takes to characterize the dehumanizers' attitude, furthermore, makes his account more rather than less compelling.

What I find problematic in Smith's position is that he does not seem to be able to indicate clear criteria attributing the combination of representations that he claims to be constitutive of the

dehumanizing attitude. Criteria for deciding whether someone attributes to someone else will, autonomy, agency, and intentions are way less problematic than criteria for deciding whether someone attributes to someone else a non-human essence. Smith directly rejects the idea that declarations to that effect are either necessary or sufficient for dehumanization but does not offer any alternative. This is a problem in the context of a theory like Smith's. It purports to be able to take into account empirical studies of human behavior, but the lack of public criteria for the attribution of some of the representations it appeals to seems to undermine this aspiration.

Furthermore, even if this problem gets somehow solved, Smith's analysis says illuminating things about the phenomenology of conduct naturally described as inhumane by connecting it with the uncanny, but in the end aims only to identify some limitations of our psychology that make the attribution of subhuman essences one of the strategies for selectively disabling inhibitions against doing violence to others. It does not aim to identify any more interesting relations between the concept of a *human being* and our conduct. Interestingly, it is conspicuously silent on what—if anything—is wrong with this strategy! As I have just pointed out, Smith describes dehumanizing attitudes as cognitively incoherent, but this does not seem to exhaust what is wrong with them. For all his theory says in this respect, they could still be overall perfectly rational attitudes to adopt in the pursuit of one's objective.

I think that humanism can and should be construed as a more ambitious idea that is compatible with the empirical and phenomenological observations made and used by Smith, but also brings into view a stronger sense in which recognizing humanity is in tension with perpetrating inhumane conduct. Humanism should be taken to confer on the notion of a human being normative, and not merely psychological, salience. In the next section I am going to articulate this alternative

way of understanding the core idea of humanism. Here I just want to draw attention to two elements that are shared by both sides of the debate as I have reconstructed it.

The first of these elements is something I have already pointed out in passing. Smith as well as Manne and the other critics of humanism all understand this view as an explanation of how certain individuals bring themselves to engage in *a certain kind of conduct*, but none of them provides a precise specification of such conduct. This is merely gestured at on the assumption that we have an intuitive grasp of what it includes.

The second element is that both sides of the debate take recognizing the humanity of the others to be a matter of representing them as having a certain feature or set of features. Smith talks about the attribution of a human essence, whereas Manne and the other critics of humanism talk about the attribution of a somewhat open-ended list of features which includes the capacity to suffer *in distinctively human ways*. What these special ways of suffering are is not articulated with any precision, but, once again, gestured at sometimes by reference to the idea of *humiliation*.

I am going to argue that the second element is an assumption that a good version of humanism should drop and that coming to see this sheds light on why the definition of the purported *explanandum* of humanism is problematic for both sides of the debate.

2.2 A DIFFERENT LINE OF DEFENSE

2.2.1 Dehumanization is Independent of Hostility

The scope of the explanation in terms of dehumanization, as ordinarily understood, is broader than conduct naturally called inhumane. In order to illustrate this point, I will consider two examples

of this kind of explanation outside of the domain with which I am mainly concerned. These are in fact contexts in which the concept of dehumanization can't possibly be spelled out as corresponding to a kind of representation of certain human beings that tends to elicit hostility or to interfere with a mechanism that normally inhibits hostility toward other human beings because, in these cases, hostility is not involved at all.

Lorraine Chuen deploys the concept of *dehumanization* in a commentary on a quantitative study of the underrepresentation of racialized people among recognized food experts who specialize in the cuisine of their own culture of origin, i.e., the underrepresentation of Mexican chefs among experts on Mexican food, Chinese chefs among experts on Chinese food, etc. She writes:

When I look at the repertoire of work that White chefs and restaurateurs have built on ethnic cuisine, it feels, in a way, dehumanizing. White people are able to establish outrageously successful careers for being experts and authorities on the stuff that racialized folks do every day simply by existing. But of course, people of colour will rarely, if ever, be called experts on how to simply be themselves. It's as if racialized folks and their ways of life are objects to be observed—study material, of sorts—rather than entire countries, cultures, and individual complex lives. (Chuen: 2017)

A second example in which the idea of dehumanization is invoked in a context that does not involve hostility is a discussion of the behavior of doctors and nurses in hospitals.

Many people report an eerie feeling about their stays in hospitals. Even if everyone treating you was kind, attentive, hard working, efficient and competent, you may still have had a sense that—compared to other situations in which people were intensely looking after you—something was different about being in the hospital. With all the measuring, palpating, listening for abnormal body sounds, injecting, and imaging of your innards, you may have felt

treated like a kind of object, rather than a complete person. You may have felt, in a word, dehumanized (Haque and Waytz 2011).

The attitudes and acts mentioned in these passages are not of the kind we would call *inhumane*. Still, they are relevant for our discussion because, in each case, the author of the passage invokes the idea of dehumanization to explain what someone finds (or might have found) wrong with them.²¹ What interests me in these cases is that they do not impute any antagonism to the agents.²² Studying a foreign cuisine is hardly a manifestation of antagonism toward its original practitioners, and taking care of sick people is, if anything is, a form of cooperative behavior.

The fact that we understand descriptions of forms of interpersonal conduct that do not involve antagonism in terms of dehumanization might seem simply to widen the area in which we tend to use a misguided explanatory schema. After all, the representations that the agents deploy in the forms of conduct in the two examples just offered (i.e. *cook from a foreign culture*, *patient*) are representations that, like *enemy* and *rival*, presuppose the attribution to the individuals for whom they are used of distinctively human features; and this is what Manne takes to be incompatible with the explanation in terms of dehumanization.

Manne argues that, though the antagonism-inducing representations she focuses on, such as *enemy* and *rival*, are occasionally used for non-human creatures as well, their primary and fullest application is to human beings. I agree with her on this point, and I think that parallel

²¹ It is not essential to my point that the judgments formulated in the two passages are true. What matters for my purposes is that they are intelligible, i.e., that they exemplify cases in which the charge of a failure to recognize certain individuals' humanity is felicitously directed at people who attribute to those very individuals distinctively human features and are not hostile to them.

²² Both passages use the notion of *dehumanizing conduct*. There are, certainly, many ways of interpreting this notion, but I think it is safe to say that, in ordinary discourse, it applies more widely than that of *inhumane conduct* precisely in that it does not entail hostility, whereas *inhumane conduct* does.

considerations hold for cooperation-inducing representations such as *patient* or *friend*. Representing other individuals according to any of these concepts presupposes, in the central cases, attributing to them distinctively human features. If we were to take the idea of dehumanization invoked in the two passages quoted above to be a matter of failing to make this attribution, we would have to conclude that the kind of confusion Manne talks about is even more pervasive than she takes it to be. It would turn out to invest not just our thought and discourse about conduct naturally characterized as inhumane, but the wider area of conduct that, though not aggressive, we find problematic in a distinctively serious way and that the two passages at the beginning of this section are meant to exemplify. But this is not the notion of dehumanization we use in ordinary contexts.

That the familiar notion of dehumanization is different emerges once we consider that, if we were to reject this sort of talk as confused, what we would end up missing is not an explanation of how some or most instances of inhumane conduct come about, but a way of capturing what makes of certain forms of conduct a single kind of conduct and, specifically, a problematic kind of conduct which calls for an explanation. The different things that we ordinarily explain in terms of dehumanization don't have in common any single feature that one might be able to appreciate independently of mastering the concepts of recognizing someone as a human being and of dehumanizing. They are not, for instance, all cases in which the inclination to attack someone is not inhibited (there might be no attack involved). They are not even cases of an agent's inflicting on someone a harm that might be well-intentioned or not-ill-intentioned, because it is not obvious we can conceptualize what happens in cases like the ones mentioned in the two passages quoted at the beginning of this section as the infliction of any harm at all without invoking the idea of dehumanization.

I take these considerations to justify the attempt to pursue a different route from the one taken by Manne, and to put into question the idea that whenever an agent applies to another individual a representation that entails the attribution of distinctively human features, she, thereby, counts as seeing that individual as a human being in the sense that matters for humanism properly understood.

2.2.2 Explanation in Moral Psychology

Manne's discussion is informed by a specific understanding of what counts as an explanation in moral psychology. This emerges from the way in which she describes a kind of conception of what seeing someone as a human being amounts to different from the one she mainly focuses on, i.e. the attribution of cooperation-eliciting capacities. This alternative kind of conception, which she only mentions in two footnotes, takes the relevant representation "human being" to have an internal relation with norms that forbid the kind of conduct naturally described as inhumane (and motivate one accordingly). In other words, according to a conception of this kind, seeing someone as a human being means seeing her as someone who has to be treated in certain ways (incompatible with inhumane conduct) and being, by the same token, motivated to do so. Manne explains why she thinks that this kind of approach is a nonstarter as follows:

If the idea of recognizing someone as a fellow human being packs in all this moral content, then it is hard to see how it could be the promised *explanans* in moral psychology. (Attributions of such recognition to an agent come precariously close to saying approvingly, "She gets it!" where the referent of "it" has been given a substantive characterization.) (Manne 2016: 404n44; see also 398n30)

What Manne finds problematic in this conception of the attitude of recognizing someone as a fellow human being is that it just stipulates the claim that is at stake between the humanist and her opponent in a disagreement where the burden of proof is on the humanist. In fact, according to Manne, there is no *prima facie* reason to think that recognizing someone as a human being should inhibit any inclination toward inhumane conduct. She thinks that any plausible version of humanism would need to work at establishing that claim (Manne 2016: 405).

Manne takes the conception just articulated to be the first horn of the dilemma faced by the humanist and focuses on what she takes to be the second one, i.e. a construal of recognizing someone as a fellow human being which does not include behavioral dispositions as part of the meaning of the relevant representation. When endowed with this kind of conception of recognizing someone as a fellow human being, humanism becomes a framework that at least purports to identify a mechanism that grounds the connection needed by the humanist explanation. The problem this time is that, as an empirical fact, this mechanism is operative only in certain specific circumstances and not in general. Manne writes:

[I]f the idea of recognizing someone as a fellow human being is thinned down to the point of being a suitable potential *explanans*, then it is not clear that it will provide the most plausible explanations of the target *explananda* all that frequently. (Manne 2016: 404-5)

We have already discussed why, by Manne's lights, a construal of seeing someone as a human being that is "thinned down," i.e. stripped of any morally substantive content, makes the familiar explanation inadequate for the paradigmatic varieties of conduct naturally described as inhumane. According to Manne, the alternatives open to the humanist therefore take the character of the two

horns of a dilemma. But is the alternative strategy that Manne rules out upfront, i.e. what she takes to be the first horn of the dilemma, really vulnerable to her objection²³?

The claim that it is depends on assuming that the point of the familiar explanation of conduct naturally characterized as inhumane is a merely causal etiology of such conduct, rather than also and, importantly, an illuminating description of some of the things that people do to each other that makes their moral significance apparent.²⁴ Manne does not take the latter alternative into consideration because she assumes that the terms that figure in the familiar explanation of inhumane conduct respectively as *explanans* and as *explanandum* are already available for thought independently of a commitment to a certain morally substantive conception of what it means to recognize someone as a human being. One might wonder where is the “moral” in Manne’s idea of explanation in “moral psychology.” Since Manne takes for granted the availability for thought of what we normally conceptualize as inhumane conduct, she thinks that dehumanization is generally treated as a factor in its causal etiology and that the task of humanism is demonstrating that it is, at least in the central cases.

But the conduct naturally characterized as inhumane is not available for thought as a single field of phenomena in a context in which a morally substantive conception of what a human being is is not already organizing the discussion. That is, inhumane conduct is not something we have any grasp of apart from our ordinary understanding of the practical significance of humanity

²³ As it will become apparent in §2.3.1, I have objections of my own to the way of pursuing this strategy that Manne considers, objections that, incidentally, undermine this strategy’s status as a form of genuine *humanism* in Manne’s own stipulated sense. To anticipate, Manne takes this strategy to rely on normative requirements internal to this or that set of human features (including e.g. rationality, capacity for humiliation, etc.) The possibility of normative requirements internal to humanity itself, for Manne, is not in view.

²⁴

without all the integrations that Manne adds to its specification in order to make of it a “natural, attractive package.” Let me explain more precisely what I mean by this.

It is worth emphasizing that, in order to *isolate* the field of phenomena whose moral psychology she is interested in, Manne has to rely on the very conceptual resources she is in the process of criticizing as inadequate to *explain* them. She introduces her topic as the moral psychology of “interpersonal conduct of the kind that is naturally described as inhumane” (Manne 2016: 390). This phrase presents the description “inhumane conduct” as somehow optional. This is not surprising because, if I understand her line of reasoning correctly, Manne would have to take the description to be misleading on the ground of a consideration that parallels the one she appeals to in criticizing the claim that the dehumanization of the victims helps to explain acts like torture and enslavement. That consideration is that an agent’s representation of the individuals she confronts as *enemies* or *rivals* can’t be in tension with the representation of the same individuals as *human beings*. Here the parallel consideration is that if we, as observers, take an agent’s conduct against certain human individuals to be intelligible in terms of the fact that she is an enemy or rival for those individuals, we cannot, at the same time, see her behavior as in tension with her own humanity in a way that makes sense of calling her conduct *inhumane*. But if we drop from the specification of the kind of conduct that constitutes the purported target of the familiar explanation the fact that it is in tension with humanity, we are left with merely antagonistic conduct, and this is much wider than our original topic.

Once one appreciates this role of the related notions of dehumanization and of inhumane conduct in the conceptualization of the kind of threats and violence characteristic of, e.g., racism and misogyny as constituting *a kind* of threats and violence, the possibilities for understanding the

explanatory significance of the appeal to dehumanization that Manne left out of the scene come back into view.

2.3 A DIFFERENT LINE OF DEFENSE FOR HUMANISM

2.3.1 Recognition

The core idea of humanism is that there is a tension, and in fact an incompatibility, between subjecting an individual to inhumane conduct and recognizing her as a human being. This recognition is not a matter of an attribution of features which causally elicits friendliness via empathy. But neither is it a matter of attributing, as *per* Manne's first horn, features like, say, rationality or capacity to experience humiliation, which are taken by philosophical stipulation to be inherently normative.²⁵²⁶ Rather it is a matter of exercising one's mastery of the first personal

²⁵ Claims with this shape, interestingly, tend to strike us as strongly revisionary unless what counts as having the relevant features for an individual is understood in such a way that no human individual turns out to lack them. Though some versions of this idea embrace the revisionary project (e.g. Singer 1974), certainly they don't constitute candidate interpretations of the content of the *familiar* explanation. Any version of this claim that aims at not being drastically revisionary will tend to resort to various ways of weakening the relation between the individual and the capacities in the set. It will go, e.g., from *possession* of the relevant (set of) feature(s) to mere *potential to develop* or *having once had the potential to develop* and the like. I think it is symptomatic of the instability of this kind of claim (as a way of capturing the content of the familiar explanation) that it really manages not to be revisionary only once the relation between the individual and the features it invokes does not do any work that would not be accomplished by saying that the individual is human—only, that is, when the appeal to the features is rendered irrelevant.

See, in this connection, Manne 2016: 396n23.

²⁶ On the criticism of the point of view according to which the only way of engaging in critical reflection on the concept of a human being is by looking for a feature or set of features possessed by human beings that justifies the way we treat them, see Diamond 1991, esp. 60-62, and Kittay 2005.

uses of the ordinary concept of a human being. The content of this concept, though not fixed once and for all by any particular piece of our tradition (whatever exactly one means by this),²⁷ is shaped quite determinately by the results of, and the ongoing contributions to, our best efforts in the activity that Diamond refers to as “making something in imagination of having a human life to lead.”

Recognizing someone as a human being is not an attitude that can be analyzed, as the Humean model would require, as being purely cognitive (or as having an identifiable cognitive component along the lines explored by Manne). This is because recognizing someone as a human being amounts to seeing an individual as being the sort of being one oneself is and as having the sort of life one oneself has, thus helping to make it the case that both oneself and the individual one confronts are, in fact, that sort of being and have that sort of life. This is because the human form of life is one that characteristically includes this mutual recognition among its bearers and much more, e.g. a claim not to be killed or enslaved by one’s fellow human beings. When I say that this mutual recognition is characteristic of the form of life we bear, I mean (among many other things) that a human being who willingly withholds recognition from other human beings is, in an important respect, wicked. A human being, independently of her individual features, is (among many other things) someone who has a claim to be recognized in this identity- and-life-shaping way. Her claim can be ignored in different kinds of ways but she remains someone with a claim to that recognition.²⁸

²⁷ It is not fixed, for instance, as Anscombe thinks it is (in the reading I am adopting), by the concepts that her version of neo-Aristotelianism retrieves from the Christian tradition (see §1.3.2).

²⁸ In making these claims, I am taking the stand described by Michael Thompson in the passage I quoted at the end of §1.3.2. That is, I am (in a defeasible but confident way) avowing for

The claims I just made are an (admittedly abstract) articulation of certain aspects of the ordinary concept of a human being that works as the background against which we exercise our capacity to understand each other and is, in fact, very basic and widely shared. It is only because we master this ordinary concept that we are able to understand why someone would say that this or that form of conduct is dehumanizing and to be sensitive to the seriousness of that claim, even when we disagree or when the form of conduct described in these terms is very far, even extremely far, from anything in our direct experience.

Primo Levi, whose memoir of Auschwitz *If This Is a Man* I have already mentioned above in a footnote as a paradigmatic example of highly aware use of the notion of dehumanization, writes:

I take the job of writing as a service to the public which has to work: the reader has to understand what I write [...] the vast majority of readers, even if they are not very prepared, have to receive my communication, I don't want to say message, which is too solemn a word, but my communication. (Levi 1986)

The description of Auschwitz as depriving its prisoners of anything recognizable as a human life certainly relies on a morally substantive understanding of what it is to have a human life to lead, but this fact hardly puts into question its value in providing an understanding of the events it is about.

In excluding what she takes to be the first horn of a dilemma faced by the humanist (Cf. p. 58), Manne argues that adopting a morally substantive conception of what recognizing someone as a human being means would be question-begging on the humanist's part. But this would be the

myself the status of a bearer of the human form whose practical understanding is sound with respect with these claims.

case only if the connection between recognizing someone as a human being and a certain way of dealing with her had to be established by the humanists in a vacuum, so to speak, and not from their standpoint as participants (in different capacities) in the various activities that contribute to the imaginative investigation of what is to have a human life to lead (for instance, listening to and telling stories, educating children, inheriting history, coming to terms with new technological possibilities).

Humanism is a situated approach. It is ironic that one might want to call “situated approach” an alternative to it according to which realizing that someone is a human being and, say, poses a danger to her leaves equally open for choice any course of action that averts that danger. I can see no other reason why someone might be tempted to propose this idea other than having succumbed to something like the Prevailing Picture. What the picture does is exactly to remove us from our position in a world whose features, such as the fact that some of its inhabitants are fellow human beings, are sources of normative requirements—in the terminology I have introduced in Chapter 1, a world in which there is a structure of value.

2.3.2 The Practical Value of the Concept of a Human Being

The lack of conceptual resources for explaining how the forms of conduct we naturally characterize as inhumane constitute a single kind might of course not be perceived as a problem if one thinks that, in fact, there is nothing significant that such forms of conduct share and that sets them apart from merely antagonistic behavior. This, however, is very implausible and, in any case, is not something that the critics of humanism I have taken into consideration mean to commit themselves to.

I want to mention, here, a further reason to think that abandoning the ordinary concept of a human being would amount to a serious conceptual loss, different from the thought that without it there would be no way of bringing together the different forms of conduct exemplified by genocidal and misogynistic threats and violence as a single kind of conduct. This further reason is connected to the discussion, addressed in § 2.2.1 above, of forms of conduct that we explain in terms of dehumanization, but that we don't call inhumane.

I want to focus on the conceptual loss involved in losing, with the ordinary concept of a human being, the connection we establish between different kinds of conduct by explaining them all as different cases of an agent's not recognizing the humanity of another individual. I am going to do so by considering an example which Manne treats as representative of the blind spot she takes to be characteristic of the advocates of humanism, namely, Cora Diamond's discussion of a passage from George Orwell's account of his experiences in the Spanish Civil War:

A man, presumably carrying a message to an officer, jumped out of the trench and ran along the top of the parapet in full view. He was half-dressed and was holding up his trousers with both hands as he ran. I refrained from shooting at him. It is true that I am a poor shot and unlikely to hit a running man at a hundred yards, and also that I was thinking chiefly about getting back to our trench while the Fascists had their attention fixed on the aeroplanes. Still, I did not shoot partly because of that detail about the trousers. I had come here to shoot at "Fascists"; but a man who is holding up his trousers isn't a "Fascist", he is visibly a fellow-creature, similar to yourself, and you don't feel like shooting at him. (Orwell 1981: 194)²⁹

Diamond defends the interpretation of the concept of a human being operative in the version of humanism that Manne excludes upfront as a nonstarter, and she uses this passage to draw attention

²⁹ Cora Diamond discusses this case in Diamond 1978: 477, and in Diamond 2001.

to the way that concept works in the practical sphere, even if Orwell writes “fellow creature” rather than “fellow human being.” In fact, Diamond comments on the passage by saying that, whether or not one would have behaved as Orwell did, one would nonetheless find intelligible the tension he points to as the same sort of tension there is in cases like this between “enemy” and “fellow human being” (Diamond 1978: 477).

As we have seen in the previous sections, Manne takes this kind of remark as a manifestation of a blind spot on the part of the humanist, because she thinks that the representation of someone as an enemy is not in tension with that of a human being but presupposes it. She comments that she does not see why one would think there is a tension between these two representations in the case described by Orwell or elsewhere, nor what might account for the fact that this tension is in place in some cases and not others (Manne 2016: 404n43).³⁰

But one can lose sight of the answers to these questions only if one looks at them through the lenses of the Humean Theory of Motivation and of a restricted understanding of what might count as explanatory in moral psychology. If one is not influenced by these assumptions, it is perfectly intelligible that any representation that, like *enemy*, has its primary application, or even just a distinctive application, in connection with human beings can enter into a tension with the representation *fellow human being*. To think of representations like *enemy* (or *friend*) as having

³⁰ The concept that Orwell represents as being in tension with that of an enemy, in Manne’s reading of his passage, is not “fellow human being,” but rather “man at his most ridiculous.” For her, what explains Orwell’s choice is this more complex representation that includes, along with humanity, something that has to do with his hierarchical position with respect to the running man and that makes it hard to see him as “fair game.” I think that this interpretation of the passage is implausible. For one thing, the detail about the trousers is quite explicitly presented by Orwell, not as something over and above the running man’s status as a fellow-creature, but as something that makes it vividly present in the circumstance. Criticizing Manne’s alternative reading of the passage directly, though, would be distracting from the main line of reasoning. I will only focus on answering the questions that she takes to be raised by Diamond’s reading.

these distinctively humanity-presupposing applications means, among other things, to think that the way and the circumstances in which one might express the stances that are associated with those representations are constrained by the way in which a human being is to be treated. Representing someone as a human being might mean, for a particular agent, among other things, that she does not see him as an enemy if he is running while holding his trousers. Constraints of this kind cannot be specified in abstraction from particular cases and it might require a considerable amount of context to grasp why one particular constraint applies in a specific case. This is why the kind of criticism of a form of conduct that we express by calling it *inhumane* or *dehumanizing* cannot be replaced by a criticism to the effect that one has contravened a certain principle of behavior: what's at issue is rather a failure of one's sensitivity.

Orwell's passage, in Diamond's reading, is particularly helpful in bringing out how this sort of sensitivity is exercised in a domain that is much wider than that having to do with thought about inhumane conduct. The kind of language that Orwell uses in connection with the prospect of shooting at the running man is not distinctive of this sort of thought.³¹

Orwell writes that he didn't *feel* like shooting at the running man, and nothing in the passage seems to suggest that he took shooting at him to be morally appalling. If this were the case, it would be strange for him to say that, in refraining from shooting, he was only partially motivated by the tension he talks about, and to provide the other reasons why he did. Clearly, thinking of shooting at a man as being, in certain circumstances, in tension with recognizing him as a fellow human being is not the same as thinking of it as being an extremely bad kind of

³¹ Diamond, for purposes related to those pursued in this section, points out that Orwell's language in this passage would not even count as "moral" according to a narrow construal of this term as equivalent to "having to do with obligations" (Diamond 2001).

wrongdoing. What is important about this reaction is that it constitutes a manifestation of the fact that our engaging with other human individuals happens against the background of a conception of what a human being is. The moral relevance of performing an action (or refraining from performing one) might depend on the fact that doing so contributes to keeping that background in place. Therefore, the conceptual resources of humanism are valuable, among other reasons, because they allow us to understand Orwell's anecdote as relevant in that way, i.e. in virtue of our understanding of what it is to engage with other human beings in the absence of that background.

The conceptual resources of humanism allow us to see very disparate forms of conduct as problematic along the same dimension even if, clearly, not in the same way.³² This, in turn, helps us to understand why an act that belongs to a practical sphere that does not have to do with obligations, and is therefore outside the domain of morality in a narrow sense, can be morally salient or part of the domain of morality conceived more broadly. Orwell's reaction to a circumstance that makes vividly present to his awareness the humanity of his political enemy is not morally salient because shooting would anyway have amounted to wronging him; it is morally salient because it is an exercise of the same sensitivity that is operative in specifically moral thought and contributes to shaping it.

2.3.3 Summary

Manne, Gopnik, and Appiah are right to say that it is very common to explain forms of conduct naturally characterized as *inhumane* in terms of dehumanization. I have argued that they are wrong

³² For the purposes of the discussion I am assuming we can agree with Orwell that shooting at the running man was morally permissible.

in maintaining that this familiar explanation embodies a confusion or blind spot. I have exploited Manne's explicit articulation of the line of thought leading to this conclusion to trace the sources of this assessment to a commitment to the Humean Theory of Motivation together with a restricted understanding of what counts as an explanation in moral psychology. Rather than attacking these assumptions directly, I have focused on the function performed by the explanation of inhumane conduct in terms of dehumanization. If this explanation is to be able (as it ordinarily is) to bring together in a single interesting field of phenomena the forms of conduct that it does bring together, the concept of a human being involved in it cannot be devoid of any morally substantive content. Contrary to Manne's contention, though, a substantive conception of what a human being is is not an *ad hoc* solution to a genuine problem faced by the philosophical view which she sees as informing the familiar explanation of inhumane conduct. It is nothing but our ordinary conception of what a human being is and, therefore, it is not question-begging for a philosophical defense of a form of explanation to rely on it.

The fact that a certain concept of a human being and the explanation of a class of phenomena it makes possible are familiar certainly doesn't make them good. But defending the value of these resources does not require one, as Manne assumes, to partition the concept into a merely cognitive representation, on the one hand, and a disposition toward a kind of conduct, on the other, and then to tell a story that justifies the combination of these ingredients in a concept. These two alleged components are not even specifiable without presupposing the ordinary concept of a human being. A criticism of this ordinary concept can only have to do with what it enables us to think, and there is this that one can say for the ordinary notion of a human being: suspending our commitment to the good standing of that concept, and with it of the ordinary explanation of inhumane conduct, amounts to depriving ourselves of the conceptual tools that make its

explanandum a topic one might think and talk about. Inhumane conduct, its specific significance with respect to other forms of wrongdoing, is something we want to be able to have in view.

Finally, I have argued that the explanation of behavior in terms of dehumanization creates connections among different regions of the practical sphere. Importantly it makes available a way to articulate what confers on e.g. racism, misogyny, and homophobia the distinctive kind of seriousness that we take them to have even in their milder and non-violent forms. Conceptualizing a remark or gesture as belonging to one or more of these attitudes amounts to taking it to manifest disregard for certain claims that the individuals on its receiving end have as recognized members of the human community and, therefore, to manifest a defect in the same responsiveness which is radically compromised in the perpetrators of inhumane conduct.

Even if Murdoch has not directly figured in my vindication, in this chapter, of the importance of the moral uses of the concept of a human being, I take such vindication to offer a particularly interesting and central example of Murdoch's idea that our moral thought does not operate in a world that we can describe independently of our moral outlook. If we deprive ourselves of the moral uses of the concept of a human being, as I have argued, we lose sight at the same time of the field of phenomena we wanted an explanation for, which is inhumane conduct and not, for instance, violent or aggressive conduct.

3.0 MURDOCH AND PRIVACY

In this chapter I focus on Murdoch's reading of Wittgenstein's treatment of privacy. As we have seen in § 1.2.1, Murdoch views Wittgenstein as having exposed the idea of inner objects as a myth, thereby revealing certain traditional philosophical problems to be the result of confusion. This achievement, though, had also helped to strengthen the position of the Prevailing Picture, not because it in fact lent any support to such a picture, but because of a misunderstanding of Wittgenstein's line of thought. The problem, according to Murdoch, is that contemporary philosophy mistakenly equates Wittgenstein's rejection of the mythical inner objects to a rejection of the existence or, at least, of the significance, of inner life. As a result, the exclusive focus on choice and action characteristic of the Prevailing Picture is presented as necessitated by logic.

The question I address here is what Murdoch's conception of privacy is and why it is so important for ethical thought.

3.1 MURDOCH'S CRITICISM OF THE GENETIC ANALYSIS

3.1.1 Privacy and Moral Progress

Murdoch schematically presents the general line of reasoning that supports the rejection of inner objects through the following two claims: (a) that such inner objects cannot be appealed to in

applying procedures for distinguishing good applications of a concept from bad ones (i.e. they are useless) and (b) that they cannot be introspectively discovered (i.e. they are not there). The second claim has been defended by appeal to both empirical and logical considerations. It has been maintained that what introspection makes available is pretty scarce and hazy and also that there are logical problems involved in the identification of such introspected materials.

Murdoch argues that Wittgenstein has only observed that a first-personal use of a mental concept verb is not a report about a private object because, lacking a checking procedure, it does not make sense to say that one is right or wrong in registering its alleged presence. The criticism of the object/report model of inner life is endorsed by Murdoch, but her position on the idea that there is no intelligible distinction between being right and taking oneself to be right when all one has is appearances is more nuanced and I will address it below (§ 3.2).³³

There is, according to Murdoch, a significant difference between, on the one hand, the observation she attributes to Wittgenstein and, on the other, an argument (the “special case” or problematic development mentioned above) various versions of which she finds in Hampshire, Hare, Ayer, Ryle, which goes beyond this observation. The problematic development, which she calls “genetic analysis” of the meaning of mental concepts, consists in deriving from the consideration that it does not make sense to take the first-personal uses of mental concepts to be reports about inner objects the conclusion that, by acquiring the capacity to apply such concepts (e.g. that of decision) in ordinary public contexts, “I learn the essence of the matter” (Murdoch

³³ As I will explain later in the chapter, this position is coherent by Murdoch’s lights because she takes procedures for determining what is going on in one’s inner life to be different from those for the presence of objects but nonetheless available (by objects I mean things that, differently from the constituents of one’s inner life, are there anyway whether or not one has any awareness of them). I address this topic in § 3.2.

1970: 12). As Murdoch immediately explains, the view she wants to put into question is that mental concepts lack any structure over and above their outer structure, which is fixed, and therefore, there is no room for the idea of progress in the understanding of them once this is grasped. There are no further steps for me to take in the understanding of a given concept once I have reached the threshold of ordinary competence in the use of the word corresponding to it. There is no transition from a concept I acquire in learning to use a word in ordinary contexts to one having to do with inner experiences of a specific sort.

I think that here Murdoch gives a specific indication of what it is that she finds implausible in the conclusions that have been drawn from Wittgenstein's work. The contrast between the genetic analysis and the conception of the meaning of mental concepts she wants to bring back into view does have to do with the opposition between purely observable circumstances and inner experiences. But the reason why this opposition matters is not just that our lives have a subjective character and it would be phenomenologically inaccurate to deny it. Privacy matters to Murdoch not just as a matter of phenomenological accuracy, but because of its connection with progress. As I will explain in the next section, Murdoch's text invites misunderstanding in this respect.

3.1.2 Two Interpretative Difficulties

Murdoch argues that the reliance on Wittgenstein's treatment of privacy to justify the genetic analysis is misguided. She characterizes the mistake as follows: "[b]ecause something is no use it has been too hastily assumed that something else isn't there" (Murdoch 1970: 10). This remark occurs at the beginning of a long discussion, and it is not immediately obvious what the "something else" is whose existence has been denied on the ground that putative inner objects have no role in the practice of concept application. What we know is that the "something else" is a phenomenon

that Murdoch illustrates through an example about a man who tries to establish, privately, whether what he feels is repentance (Murdoch 1970: 25), to whom I will refer as R, and through the widely discussed example of M and D (Murdoch 1970: 16-17). I will discuss the second example at length both in this chapter and in the next, in connection with a different problem. The only feature that matters for my purposes in this section is that each of these cases, in Murdoch's intentions, constitutes an attempt (successful in M's case) at a special kind of moral progress which is connected with the idea of perfection.

Before discussing these examples exploited by Murdoch, I need to point out two difficulties in Murdoch's discussion of privacy. The first difficulty is why she chooses this kind of example.

We can start with the formulation, just quoted, of the mistake she identifies in the genetic argument. If I read Murdoch correctly, the thing that has been demonstrated to be useless and the different thing which has been mistakenly taken to be not there are to be identified as follows.

What we can come to recognize as useless thanks to Wittgenstein's work in this area are inner objects which are thought to be just like any other object but have the peculiarity of being cognitively available only to one subject (the one whose interiority they belong to). Furthermore, somehow, this subject is infallible in this cognitive feat, as if she could look into a walled garden (according to the image used by Crispin Wright 1989) where such objects lie. Finally, these objects are the realities to which mental concepts correspond: their presence is what makes the application of mental concepts correct.

What has been too hastily concluded to be absent is inner life *under any understanding of it*. This entails the exclusion of what we ordinarily, i.e. without being driven by philosophical worries, would consider our inner life, which is what provides, for instance, answers to questions like "what's on your mind?" Inner life in this yet to be specified sense of course includes much

more than the sort of circumstances that makes it the case, for instance, that a man might rightly judge, after long and difficult private reflection, that what he feels is after all repentance. It includes, for instance, completely mundane and uninteresting inner experiences, such as remembering, at one particular moment, that one has an errand to run.

For this reason, I think that there is something that calls for an explanation. This is the fact that Murdoch focuses on cases that she takes to exemplify a distinctive and important kind of moral progress connected with the idea of perfection. Why consider these cases if those about remembering about an errand would work just as well to reject the genetic analysis, i.e. a behaviorist reading of Wittgenstein's treatment of privacy?

A hypothesis is that her choice of cases depends on the particular themes of "The Idea of Perfection," which is motivated specifically by her interest in the philosophical understanding of moral personality. Questions about repentance are more typically connected with morality than the ordinary operations of our memory are. According to this way of understanding Murdoch's chosen strategy, she might have just as well chosen to criticize the genetic argument by talking about any element of our mental life and not simply those that have moral relevance. In choosing the examples as she did, though, she showed not only that life which is, in a certain specified sense, private is there, but also that some of its episodes are morally salient. Whereas treating the inner experiences of remembering that one has an errand to run is pretty inconsequential—what matters is, after all, whether one runs the errand—it is not inconsequential to do so with inner experiences that count as moral achievements or, at any rate, inner experiences that Murdoch invites us to recognize as such.

If this were the case Murdoch could have rejected the behaviorist interpretation of Wittgenstein's treatment of privacy by arguing, as others have done, that accepting it is compatible with the following claim:

It is not silent and dark within: We can understand instantiations of mental concepts as mental states brought about by mental processes which are not shadows of observable behavior in the sense that they obtain whether or not they find an expression in behavior.

The sentence "it is not silent and dark within" is used by Murdoch (1970: 13) in the course of her argument, and in what follows I will use it as a shorthand for this more articulated and less metaphorical claim. In order to establish this claim, one certainly does not need to rely on the intelligibility of cases in which the unfolding of the inner life of the subject counts as moral progress, let alone moral progress of a special kind which is connected with the idea of perfection. Yet, according to the interpretative hypothesis I am considering, once the claim that it is not silent and dark within is rescued from the behaviorist attack, one has all that is needed to have the sort of episodes Murdoch is interested in back on the scene, and she might then go on to discuss their moral relevance. Their moral relevance, though, would be quite independent from the fact that they are episodes in inner life. In other words, it would not be the fact that they are private that makes them morally relevant: it might be not silent and dark within, and still very boring.

Certainly, there is no reason to saddle Murdoch with the view that any application of a mental concept is necessarily governed by the effort to strive toward perfection. Her view is compatible not only with the claim that sometimes all there is to a first personal mental concept application is the obtaining of certain publicly observable circumstances (as she explicitly concedes in Murdoch 1970: 14), but also with the claim that pretty mundane inner experiences, such as that of suddenly remembering one has an errand to run, are part of one's inner life too.

Nonetheless she repeatedly suggests that the sort of cases that she invites us to recognize as involving a striving toward perfection are needed for her attack to the genetic argument.³⁴ She writes, for instance:

As soon as we begin to use words such as ‘love’ and ‘justice’ in characterizing M, we introduce into the whole conceptual picture of her situation the idea of progress, that is the idea of perfection: and it is just the presence of this idea which demands an analysis of mental concepts which is different from the genetic one. (Murdoch 1970: 23).

Let me summarize what I take to be the first interpretative difficulty about Murdoch’s criticism of the genetic analysis of mental concepts, that is, the idea that their meaning is exhausted by the observable circumstances in which one learns how to apply them in public contexts. It seems that, if we can vindicate the claim that it is not silent and dark within, then we’ll have thereby refuted the genetic analysis of mental concepts. But, in order to do so, it is sufficient to show that episodes, such as remembering that we have an errand to run, which are not morally relevant can, in fact, be understood as belonging to inner life. Nonetheless, Murdoch thinks that episodes of inner life that exemplify a kind of moral progress connected with the idea of perfection are essential to her argument.

Of course, one might think that this is a mistake on Murdoch’s part. But what I take to be the case is that what can be imputed to her is just a confusing exposition of her argument. More specifically, I find her argument confusing in that she fails to clearly separate two different defects

³⁴ The claim that the inadequacy of the genetic analysis is connected with the idea of perfection and with the fact that human beings are historical individual is something that Murdoch repeats frequently (see esp. Murdoch 1970: 23-28).

of the genetic analysis: (1) a behavioristic understanding of the circumstances that license the first-personal application of a mental concept and (2) a less specific and weaker idea of the identity of mental occurrences as dependent on public rules. While to reject (1) it is enough to bring out the perfectly non-mythical sense in which a first personal application of a mental concepts can be in order in the absence of publicly observable circumstances that license it, to destroy (2) it is necessary to bring back into view a different sense in which concept applications can be independent from public rules.

This different sense marks the kind of privacy proper to the episodes connected with the special kind of moral progress Murdoch wants to protect, which is different from the kind of privacy proper to episodes such as remembering that one has an errand to run. I am going to explain what this independence is and how it is illustrated by the examples chosen by Murdoch in the next section, but I will introduce the second interpretative difficulty first.

I have just said that there is a sort of privacy which consists in an independence from public rules of meaning, and which is stronger than the one entailed by the claim that it is not silent and dark within. This privacy is a result of the connection between a concept and a kind of moral progress involving the idea of perfection. Murdoch repeats several times that mental concepts have this connection (e.g. 1970: 23-24 and 28) but she also says that perhaps all concepts are connected in this way with the idea of perfection (1970: 29) and, as I have already mentioned, concedes that there are uses of mental concepts that do not involve the idea of perfection at all (1970: 14-15, 35). Most importantly, Murdoch does not provide a characterization of what counts as a mental concept.

I think that the best way to understand what, according to Murdoch, is connected with the idea of perfection is in terms of what she calls “specialized uses of concepts” (1970: 25) which are

necessarily private. Though mental concepts paradigmatically lend themselves to such uses, any concept might end up being used in this way. Of course, it remains to be explained what specialized uses of concepts are, how they can be meaningful despite their privacy, and what their distinctive privacy has to do with moral significance. I will address these questions in the next section (§ 3.2) by looking at one of the features of the examples offered by Murdoch to challenge the genetic analysis. In the next subsection (§ 3.1.3) I will introduce the examples and discuss two further features of them.

3.1.3 Is Murdoch's Conception of Privacy Mythical?³⁵

As I have anticipated, the two main examples that Murdoch uses to illustrate her conception of privacy are the example of M and D and the example of R, "a man trying privately to determine whether something which he 'feels' is repentance or not" (1970: 25).

The role of the example of M and D in Murdoch's argumentation is complex because it serves many different purposes. In this subsection, I will summarize it and I will identify three features of the activity which is described in it, which are shared by the example about repentance. I think that while the first two features have to do with the claim that it is not silent and dark within, i.e. that accepting Wittgenstein's criticism of privacy does not force us to give up the very idea of interiority, the third one goes beyond that claim and its associated idea of privacy. It is this third feature that brings out the concept of Murdochian privacy which is connected with moral progress and I will discuss it separately in the next section (§ 3.2)

³⁵ In this subsection and the following ones in the chapter I draw on McDowell 1998 and McDowell 2009 for the interpretation of Wittgenstein's treatment of privacy.

Here is the example of M and D: D is a young woman and M is her mother-in-law. M experiences a moral progress that consists in transforming her way of seeing D through an effort of self-criticism and attention toward her. M considers the possibility that she herself is “snobbish”, “conventional”, “narrow-minded” and admits to herself that she is “jealous”. This self-criticism leads her to reconsider her idea of D and to engage in an effort of attention toward D, as a result of which she goes from thinking that D is “unpolished,” “lacking in dignity and refinement,” “tiresomely juvenile” to thinking that she is “spontaneous,” “gay,” “delightfully youthful” (Murdoch 1970: 16-17). Despite the radical change in her way of seeing D, M had behaved perfectly kindly toward D all along, so M’s transformation in the way she sees D does not result in any change in her behavior.

Murdoch characterizes the kind of phenomenon that these two cases are meant to exemplify (1) as a specific kind of activity, (2) as something that is not hazy, but rather something that we find very familiar and (3) something that is also essentially one’s own, something that could not be done in conversation with someone else, but not for that reason infallible.

Let me now examine the first two elements that differentiate Murdoch’s characterization of inner life from the picture of interiority rejected by Wittgenstein.

The first is Murdoch’s view on how the notion of activity has to be understood in connection with inner life. She writes:

I am now inclined to think that it is pointless, when confronted with the existentialist-behaviorist picture of the mind, to go on endlessly fretting about the identification of particular inner events, and attempting to defend an account of M as ‘active’ by producing, as it were, a series of indubitably objective little things. “Not a report” need not entail “not an activity.” (Murdoch 1970: 23)

Murdoch's point here is that her example of M is not meant to rescue the intelligibility of some version of the idea of inner data. But the fact that there are no data to report about does not exclude that there is *activity* inside. From the perspective of the advocate of the genetic analysis, to be sure, this does not bring about much progress. The problem with what is inner, from this perspective, has nothing specifically to do with the categories in terms of which one characterizes it, but with the absence of public criteria that can establish its presence/occurrence inside a person. Objects and activities are on a par in this respect. Murdoch's defense of inner life has to do with the specific *kind* of activity she takes it to be. The main point is that inner life is something we engage in, not something we witness; what goes on inside does not go on independently from our awareness of it, and it is precisely this idea of "going on independently of our awareness" that is implicitly suggested by the characterization of the alleged inner data as "objective," and as something one can "report" about. But Murdoch explicitly distances herself from an understanding of M's activity as reducible to a complex of such data. Quite to the contrary, M's example brings to the fore in an especially vivid way that vision of how things are is not a given but an achievement. In it, the involvement of the subject's will takes the shape of a deliberate effort. We are very far from the imagery of the walled garden to which one has direct access.

Let me emphasize here that, at this stage of the discussion, the mental concept raising a problem for the genetic analysis of its meaning is "change of mind." The concepts that M uses in her activity are not yet relevant for the argumentation. It is the concept of change of mind that the example shows to be applicable in the absence of any observable circumstance. The notion of activity invoked by Murdoch has to do with rejecting the claim that allowing for this possibility turns its instantiations into mythical inner objects. We fall into the mythical conception of mental states and processes if we take them to be states and processes that are there independently of our

awareness of them. But we do not fall into this mythical conception if we acknowledge that these states and processes, as McDowell has put it, “have no being independently of the fact that the concepts they instantiate figure in the content of consciousness” (1998: 311). All there is to M’s knowing that she has changed her mind is precisely her having *done* so, which is something that one does consciously.

It might be worth emphasizing that Wittgenstein makes room for the idea of what might be broadly called “the same sort of happenings” in the mind, but in connection with applications of mental concepts that have nothing specifically *moral* about them. For instance, in PI §662, he writes about someone who beckons another person, called N., in the context of a series of passages that exploit the idea of appropriately used signals as a model for applications of mental concepts. The model is meant to emphasize the internal relations between the instantiation of the mental concept and its verbal expression—relations that might be obscured by thinking of such a verbal expression as a report. Wittgenstein writes: “One can now say that the words ‘I wanted N. to come to me’ describe the state of my mind at that time; and again one may *not* say so.” The reason why one might not say so is precisely the risk that these words encourage the idea of the problematic independence of the mental occurrence from one’s awareness of it. Of course, this raises the issue that, if the instantiations of mental concepts are not independent of the subject’s awareness of them, there is no room for the subject to make a distinction between being correct and seeming to be correct in her first-personal uses of these concepts. This is, for my purposes, the most important point and I will come back to it at the end of the next subsection, in my discussion of the third feature of the examples used by Murdoch. As I have anticipated, it is this third feature that brings out the concept of Murdochean privacy, which is connected with moral progress and which goes beyond the scope of vindicating the thought that it is not silent and dark within.

The second element of Murdoch's characterization of inner life (which is related to the first) is that if we drop the idea that the only thing inner life could be is some sort of parade of inner items for our inner eye to contemplate, the inner scene turns out to be way less elusive than that model leads us to think. There is nothing hazy (dubious) in our ordinary self-ascriptions of thoughts, intentions, wantings and the like: they simply are exercises of our capacity to apply these concepts. Murdoch brings this out by drawing our attention to how familiar and non-mysterious an activity like M's is, and how similar it is to the sort of activity that is often described in novels and that we find completely intelligible (Murdoch 1970: 22; see also Murdoch 1951: 30-31).

Once again, we find this point expressed by Wittgenstein in connection with more mundane instances of inner life, like remembering that one would have finished a sentence in a certain way, had she not been interrupted. I follow McDowell (2009: 88-90) in reading PI §§633-635 as showing that the task of completing what one was going to say before an interruption seems to be one of "reading the darkness," in Wittgenstein's view, only when one looks for elements that would *explain* her claim that she would have indeed finished the sentence in a particular way, rather than to take her to be exercising her competence in the relevant practice—by saying, simply, that such and so is, in fact, what she had meant to say. An appeal to that intention would not count as individuating a reality that justifies the way we talk; it would be just one more instance of talking that way, and therefore would not count as an explanation of the kind sought for in the rejected interpretation of the task. But the need for such an explanation will not be felt, or even thought of as intelligible, once the character of mental concepts is properly understood.

What emerges from Wittgenstein's discussion of the problems with the mythical conception of inner objects is not that there really are no mental states and processes after all, i.e., that so called mental states are really nothing but shadows of publicly observable behaviors.

Rather, the lesson to be drawn is that, when we use mental verbs in the first person, we are neither latching onto “indubitably objective little things” nor surreptitiously referring to patterns of external circumstances. The understanding of what a mental concept is reveals that there need be nothing in the correctness of a first personal use of a mental concept apart from its subject’s competence in the practice in which that concept has a role. This however, clearly does not imply that the subject has to wait and observe her own subsequent behavior in order to attribute to herself, for instance, a determinate intention (as the shadow view would have it). A subject knows her own intentions, in their full determinacy, by forming them. Her competence with the use of the concept of intention does not contribute to identifying the intention by being exercised in overt behavior, but in being presupposed in the very self-ascription of that intention.

Here, though, I need to note an important difference between the examples of inner life exploited by Murdoch and the more mundane example used by Wittgenstein, and by McDowell in interpreting him, that I have just mentioned. M’s activity, as Murdoch emphasizes, is of a familiar kind, but it is difficult. Granted, it is not difficult in a way that has to do with the weird attempt to register, or retrieve from memory, “purely inner data,” and yet it involves more than basic competence in the linguistic practices involving the concepts she uses in it. As Murdoch says: “M’s activity is hard to characterize not because it is hazy, but *precisely because it is moral*” (Murdoch 1970: 22). I’ll explain what I take Murdoch to mean by this claim in my discussion of the last of my three points.

3.2 IDIOSYNCRASY AND THE IDEA OF PERFECTION

3.2.1 Idiosyncrasy

The third feature of the inner activity illustrated by Murdoch's examples of M and D and of R is, as I have anticipated above, that it is something that is essentially one's own, something that could not be done in conversation with someone else, but not for that reason infallible.

This is an observation that reclaims a notion of privacy different from the one that the picture targeted by Wittgenstein tries and fails to make sense of, but also from that implied by the mere fact that we do have inner lives. The intelligibility and importance of this further kind of privacy is a distinctive feature of Murdoch's attack on the genetic analysis of mental concepts which differentiates it from the one we find, e.g., in McDowell.

In "Intentionality and Interiority in Wittgenstein", McDowell—immediately after articulating the thought that a non-mythical conception of inner life requires us to recognize that its episodes do not take place independently of their subject's awareness of them—addresses the worry that the acknowledgement of this dependence might be perceived as impugning their reality. He explains:

It is true that no distinction between "seems right" and "is right" opens up, with respect to the obtaining of these states of affairs, from the subject's point of view. But what it is for the relevant concepts—those whose instantiation in consciousness exhausts the being of those states of affairs—to have application cannot be understood exclusively from the subject's point of view. The concepts set up internal links between the states of affairs that are their instantiations and publicly accessible circumstances linked "normatively" to the states of affairs in one kind of case [that of intentional mental states], circumstances linked to them as their normal expression in another [that of sensations]. So the distinction between "seems right" and "is right" opens up from the necessarily thinkable second-person or

third-person point of view; one's own inner world is part of the world (anyone's world). (McDowell 1998: 311)

This passage retrieves the distinction between “seems right” and “is right” with respect to the obtaining of mental states by exploiting (1) publicly accessible circumstances that help to fix the meaning of the concepts for both intentional mental states and non-intentional ones and (2) the thinkability of such mental states in the second and third person perspective. But Murdoch's objective is to show the need to radically rethink the role of both these resources in the light of the existence of the specialized uses of concepts.

The sort of attack on the genetic analysis that we find in McDowell rejects the idea that our states of mind have a parasitical (or “shadowy”) existence. There is no need for any publicly observable performance to obtain for them to be intelligibly and determinately present in our interiority. But the fact that one might enter a mental state depends on something public, namely, her competence in the practice in which the corresponding concept has a role, and this is all that is needed to make her self-ascription authoritative. McDowell uses the example of the intention to type a period (1998: 315). Certainly, he argues, this intention could be fully formed in a subject and could be correctly self-ascribed independently of any subsequent performance. But the possibility of forming it depends on the fact that she is “party to the practices that are constitutive of the relevant concepts” (1998: 315). In this sense, the unactualized intention is a mental state whose existence is not parasitical and yet it is not mythical precisely because of its connection with certain public practices which include the ascription of intentions to others.

This is all perfectly satisfying with respect to mental states such as the intention to type a period. Considerations along these lines can also be said to hold across the board. The link between a subject's first personal uses of certain mental concepts and her competence in the practice of

applying them in other ways is required if her inner world is to be part of the world. But Murdoch's examples present us with reasons for thinking that this cannot be the entire story.

She draws our attention to the fact that the activity of reassessing and redefining one's mental concepts and the way one applies them "often suggests and demands a checking procedure which is a function of an individual history" (1970: 25). In other words, this checking procedure involves precisely those aspects of our thought that go beyond what we share with most other people and in which we use concepts in a way that is shaped by our personal history, i.e., in ways that are "specialized" in her sense of the term.

I want to highlight that the very idea of "specialization" presupposes a basis one builds upon. A specialized use of a concept, in Murdoch's conception, can be highly personal. But if it makes sense to call it a specialized use of the concept of, say, repentance such use must come from someone who has competence with its ordinary and conventional use. The independence from public criteria of meaning that the idiosyncrasy of specialized uses involves does not entail that such criteria are dispensed with. It entails that they are unable to settle certain important questions about their proper application and that second and third-personal uses of such concepts, even if thinkable, might not be really ever at stake.

The examples of M and D and the example of R, then, exhibit a kind of privacy that does not derive from the mere interiority of their activity. This is a feature that this activity shares with unexecuted (and un-manifested) intentions to type a period. They exhibit a privacy that attaches not to the concepts of change of mind, in the case of M, and of self-scrutiny, in the case of R, but to the concepts that each of the two protagonists is trying to apply properly in a sense that requires more than mere ordinary competence. Ordinary competence is part of what each character needs

in order to even engage in his or her respective task, but it is far from sufficient to bring it about.

Murdoch writes:

There are two senses of “knowing the meaning of a word,” one connected with ordinary language and the other much less so. (Murdoch 1970: 28)

This should not strike us as mysterious, since the two senses are not completely separate. As I said, the term “specialization” suggests that the second includes the first and therefore there is no risk here of losing the link with public practices. Furthermore, that concepts lend themselves to specialized uses is a very familiar fact of life. Even if we do acquire our concepts in public contexts, some of them, and particularly mental ones, characteristically undergo a process of transformation. And here we reach the connection between the idiosyncrasy of specialized uses of concepts and moral progress.

The way in which we understand concepts like courage and repentance, in fact, is not the same in different phases of life. We are invited or pressured to change our concepts, or the way we apply them, by the specific objects of attention that are our own. In fact, there is something wrong with someone who does not go through this process of reassessing. Murdoch calls this process “the main characteristic of *live* personality” (1970: 25, emphasis added).

Of course, the transformation of our concepts can be a case of degeneration and corruption, and I will come back to this problem in the next subsection. But, for now, the point is that the historical nature of human individuals brings in a perfectly intelligible and non-mysterious sense in which our inner lives, where these transformations take place, are private, and they lead to concepts that are, therefore, idiosyncratic. Nobody shares our history. The acknowledgment of

idiosyncrasy entails no more than the thought that what one really means in applying a mental concept can only be understood under certain conditions.

Human beings are obscure to each other, in certain respects which are particularly relevant to morality, unless they are mutual objects of attention or have common objects of attention, since this affects the degree of elaboration of a common vocabulary. (Murdoch 1970: 32)

The privacy of inner life Murdoch works to vindicate goes beyond the fact that sometimes we do not express our mental states in behavior. It is a consequence of the historical nature of human individuals, which makes it the case that in many cases we would not be able to communicate them to others. The reason is not that our mental states are inexpressible in principle. In the passage I have just quoted Murdoch even gives a very quick and yet precise characterization of the circumstances in which the elaboration of a common vocabulary is possible. But to say that this communicability is possible in principle should not, so to speak, domesticate the point. The fact that a shared world is only a remote ideal is a structural aspect of our condition. Murdoch writes:

that “reality” that we are so naturally led to think of as what is revealed by just “attention”, can of course, given the variety of human personality and situation, only be thought as “one”, as a single object for all men, in some very remote and ideal sense. (Murdoch 1970: 37)

The passage just quoted introduces the different standard for good application proper to specialized uses of concepts. Murdoch’s notion of idiosyncrasy, in fact, does not involve giving up on the need for a checking procedure. This is the topic of the next subsection.

3.2.2 Imagination and Fantasy

I have already discussed in Chapter 1 Murdoch's move to bring back into view the infinitely elusive character of reality. The fact that the description of the world produced by science and common sense serves important human interests does not require us to treat it as the ultimate description or as setting the standards of factuality. And once this purported requirement is dropped, the way in which we experience the world as moral beings can be retrieved as something that moral philosophy does not simply allow for but has to take into account.

The concepts that we need and that we use in our ordinary lives are not those of science and common sense with the addition of a limited set of "moral words" connected with choice and freedom. Our vocabularies and the ways we use them have a characteristic tendency to become specialized in response to the specificity of the particular sets of experiences that make up our individual histories, unless of course we are utterly conventional—a defect that, in Murdoch's framework, can be characterized as completely failing to be sensitive to the infinitely elusive character of reality.³⁶

In the previous subsection I have argued that Murdoch's appeal to certain familiar experiences (e.g. the way in which the concept of courage changes during the life of a human being) and to a pair of examples reminds us of the fact that some of our concepts, especially mental concepts, change and become specialized without thereby becoming unintelligible. I said that questions about the criteria of correctness for specialized uses of concepts cannot be settled with reference to public criteria of competence, because for a concept to acquire a specialized use is

³⁶ Murdoch writes, for instance: "It is characteristic of morals that one cannot rest entirely at the conventional level, and that in some ways one ought not to" (Murdoch 1970: 29).

precisely for it to depart in important respects from its public meaning. The question, then, is what makes a specialized use of a concept correct. The short answer is, its capacity to disclose reality to the subject.

This admittedly tautological answer becomes more informative if we consider, so to speak, two poles, two points that set opposite directions for the development of our concepts. Murdoch argues that our uses of concepts can become more personal and idiosyncratic as a result of our getting trapped in fantasies or as a result of our use of imagination. “Fantasy” and “imagination” name, respectively, the poles that sets the direction of degeneration and the direction of progress in the development of attempts to get to reach reality. An alternative conceptual pair that Murdoch uses to identify to these two poles is selfishness and love or justice.

It is very interesting that Murdoch chooses, in order to characterize M’s idea of D, at the beginning and the end of the story, respectively, concepts whose public criteria of application are the same (“tiresomely juvenile”-“delightfully youthful”; “vulgar”-“refreshingly simple”). What M needs to face the situation she finds herself in is not a better command of these criteria, but an effort of attention. Murdoch’s description of the example portrays M as able to identify specific obstacles to the proper appreciation of reality. M considers the possibility that she is “conventional” and “old-fashioned,” and these are two ways of lacking the capacity or willingness to change one’s way of looking at things. M also admits to herself that she is jealous, and this is a selfish emotion. The opportunity for change is provided, in the example, by M’s newly found capacity to consider the otherness of D. It is by turning her focus away from her own habits, needs, and wishes and to D that M grows and changes for the better. This is, in the first place, what Murdoch asks us to find plausible and persuasive.

What we see in the example of M and D then is a situation that calls for specialized uses of concepts and a subject who succeeds in changing her way of reading her situation by overcoming certain forms of conventionality (snobbish, old-fashioned habits of thought) and of selfishness (jealousy) by turning her attention to an external reality (D).

It might be said that R, whose situation Murdoch does not describe in detail, faces the same sort of task. If we try to make the situation, vivid to ourselves we can easily imagine or remember being in a similar situation and it seems plausible that the obstacles that hinder the resolution of the sort of doubt that troubles R are again convention and selfishness. R might be just frustrated because of the consequences of what he has done or have somewhat automatically assumed a contrite attitude in reaction to other people's expectations. It is hard to imagine what going about R's investigation might possibly mean if not checking on these possibilities. Doing so typically requires looking outside oneself not at external circumstances that would settle the matter (they can't) but at relevant aspects of the situation, for instance, the pain one has caused.

The identification, on Murdoch's part, of these two particular obstacles, conventionality and selfishness, as the two sources of our failure to see reality for what it is depends on the conception of reality as infinitely elusive and a broadly Freudian conception of a human being. As I have mentioned above, the conception of reality as infinitely elusive requires that we develop through imagination specialized uses of concepts, and being conventional is precisely being incapable of or unwilling to do so. I now turn to selfishness which, though not incompatible with conventionality, characteristically results in the development of specialized uses of concepts that tend to distort reality rather than revealing it in more detail.

About the lesson she draws from Freud, Murdoch writes:

What Freud presents us with is a realistic and detailed picture of the fallen man [...] He sees the psyche as an egocentric system of quasi-mechanical energy, largely determined by its own individual history, whose natural attachments are sexual, ambiguous, and hard for the subject to understand and control. (Murdoch 1970: 50)

Though Murdoch claims that she is not a Freudian and is, as we have seen, very critical of certain uses of psychoanalysis (cf. §1.2.3), she finds in Freud's works a persuasive representation of human beings. In particular, she finds there an understanding of the ego as the main source of obstacles to the appreciation of reality, in virtue of its capacity to generate fantasies out of our fears, obsessions, and traumas.

In the description of the two main examples she exploits, Murdoch presents us with a very realistic representation of a kind of moral problem and of what counts as making progress in solving it and what the resources for doing so are.

At a higher level of abstraction, Murdoch describes M (but, given her understanding of his task, she would characterize R in these terms as well) as trying to look at D “justly” or “lovingly” and some people might want to resist the second of these two characterizations (“lovingly”) for various reasons. But nothing important depends on this word-choice on Murdoch's part. The attitude that she invokes as relevant to determining whether a specialized use of a concept is a case of progress is to be understood, primarily, as the attitude consisting in turning toward reality by overcoming a certain very specific sort of obstacles: convention and selfishness. The fact that not everyone would find it satisfying to call this attitude “love” depends on the fact that “love” is a concept that is virtually impossible not to specialize and Murdoch herself makes a specialized use of this concept. I turn to an articulation of this concept in the next chapter. Before doing so I want to make a last consideration about the checking procedure for specialized uses of concepts.

The question about the checking procedure requires two slightly different but strictly connected answers in the case of mental concepts, which I understand as those whose instantiations are exhausted by our awareness of them, and the others. The case of R involves a specialized use of a concept of the first kind, and that of M and D mostly specialized uses of concepts of the second kind. Both M's and R's specialized uses of concepts count as correct if M and R display, apart from ordinary competence in using these concepts in a non-specialized way, a high degree of sensitivity to the specificity of their situations, which in turn requires a certain capacity to resist convention and selfishness. But M's use of the concept "refreshingly simple," if it is correct, might also have been arrived at by someone else and not only by M. Certainly, this other person would have to be in a position very similar to M's, but it is not unthinkable, just very unlikely. On the other hand, if R forms a judgment to the effect that, say, what he feels is really repentance and this judgment is correct (is performed justly and lovingly), it is not the case that it could have been arrived at by someone else and not by R.

If one accepts, as I think one should, the understanding of mental states as such according to which their being is exhausted by their instantiation in consciousness, one would have to re-describe R's situation as one in which a concept of repentance is instantiated, so that his doubt is really about whether his attitude is really just and loving and not about whether the instantiation occurs or not.

4.0 MURDOCH AND LOVE

Murdoch often specifies her own view in moral philosophy in terms of a contrast either with Kant or with a complex of ideas that she represents as resulting largely from Kant's influence on contemporary thought. It has been occasionally maintained, though, that Murdoch overestimates the extent of her disagreement with Kant and that her distinctive theses can be integrated into, or reformulated in terms of, the Kantian framework. In this chapter, I rely on the distinction I have drawn in the line of argumentation in *The Sovereignty of Good* between two levels, dealing with two different targets, to achieve a proper appreciation of Murdoch's distance from Kant and to bring into view what I take to be the advantages of Murdoch's proposal. Despite certain elements of commonality that Murdoch was, in fact, ready to acknowledge, there is a fundamental difference in the role and in the character of the two attitudes that Kant and Murdoch, respectively, indicate as the distinctively moral one: respect and loving attention.

4.1 TEMPERAMENTS AND STRATEGIES

In her criticism of contemporary moral philosophy, Murdoch is concerned with recovering the possibility of assessing genuinely alternative moral theories. This possibility, in her opinion, has been precluded by a widespread and unwarranted acceptance of what I have been referring to as "the Prevailing Picture." As we have seen, this picture amounts to a complex of ideas within which Kant's influence operates in combination with other forces, and this entails significant distortions. The Prevailing Picture, in other words, is Kantian only in a thin sense and it is to be distinguished

from the moral theory that Murdoch attributes to Kant. At the first of the two levels of her line of argument, Murdoch criticizes the presuppositions that make the Prevailing Picture seem inescapable, and, at the second, she presents her own view as a rival to both the Prevailing Picture and the moral theory she attributes to Kant. This chapter is mostly concerned with the second of these two levels.

Even if Murdoch's judgment of the neo-Kantianism of the Prevailing Picture is much harsher than her judgment of Kant's moral philosophy, Murdoch takes her own view to be different from and superior to both. Some of the criticisms that Murdoch levels against the Kantianism operative within the Prevailing Picture, as she explicitly points out, do not apply to Kant's moral philosophy, but this should not make us lose sight of the fact that her proposed view remains, nonetheless, radically alternative to the latter as well. I will explain why this is so by criticizing two recent attempts to interpret Murdoch's work in moral philosophy as compatible with Kant's framework. The first is David Velleman's claim that Murdoch's conception of love, rather than challenging the adequacy of Kantian ethics, actually provides the resources for answering an objection to it (Velleman 1999). The second is Carla Bagnoli's claim that Murdoch was ultimately wrong in describing her view as an alternative to Kant's since "Kant and Murdoch share the same view of moral sensitivity and agency" (Bagnoli 2003: 510).

As we have seen in § 1.2.2, Murdoch protests against certain specific uses of the Kantian idea of the world we inhabit as, *per se*, devoid of any value and of the individual as generating value through the operations of her will rather than apprehending it. These uses result in a picture of moral thought as unambitious and unrealistic. This kind of appropriation of Kant's work, which Murdoch associates with Kantian liberalism is in fact interested in reasons but, in its emphasis on their public and impersonal character, makes of moral reasoning a trivial matter of surveying facts

and applying principles. What goes missing in this approach is an aspect, indeed *the* characterizing aspect, of our situation as moral beings, i.e. the fact that the appreciation of the facts that constitute our circumstances is itself extremely difficult.

The only facts Kantian liberalism envisions are ordinary public facts that become action-guiding only when they are elected as reasons by the unconstrained, but also unguided, will; and the only meanings it admits are ordinary meanings fixed by public criteria. Any possible idiosyncrasy in the appreciation of facts or in the use of words deriving from the history of the individual who is the subject of that appreciation and use is eliminable, in principle, by psychoanalysis, conceived of as a discipline specifically devoted to reabsorbing into the sphere of quasi-scientific objectivity anything that might have struck us as personal and private in the relation between the individual and its circumstances. The connections that make the world meaningful and salient for us are eliminated from the representation of our position to the point that the very idea of reasons for choosing anything over its alternative loses meaning. I think that the idea is brilliantly illustrated by Murdoch's device of representing neo-Kantianism (in her sense) and Surrealism as the two wings of the existentialist Prevailing Picture, where the distance between them is marked only by the degree of interest in, or perhaps we should say trust in the intelligibility of, reasons, which is strong for the neo-Kantian and completely absent in the Surrealist (Murdoch 1970: 34).

Kantian liberalism, in Murdoch's reading, does not provide an ideal limit to strive toward, but rather something like a minimum threshold. This, for her, amounts to being inadequate according to both criteria that she explicitly offers for the assessment of moral theories. In order to be satisfactory, she proposes, a moral theory must in the first place be "realistic," that is, it must take into account the features of concrete human beings. In the second place, since no moral theory

can be neutral, it has to orient people's efforts to become better by conveying a "worthy ideal" (Murdoch 1970: 76).

Murdoch does not extend this kind of criticism of Kantian liberalism to Kant himself. She thinks that Kant's emphasis on the purity of the will and on notions such as responsibility and freedom might encourage an interpretation of Kant as an existentialist in her sense—someone who identifies the will as the only source of value in an otherwise value-free reality—but this, she thinks, would be a misunderstanding. She writes:

It must be remembered that Kant was a 'metaphysical naturalist' and not an existentialist. Reason itself is for him an ideal limit: indeed his term 'Idea of Reason' expresses precisely that endless aspiration to perfection which is characteristic of moral activity. His is not the 'achieved' or 'given' reason which belongs to 'ordinary language' and convention, nor is his man on the other hand totally unguided and alone. There exists a moral reality, a real though infinitely distant standard: the difficulties of understanding and imitating remain. (Murdoch 1970: 30)

Murdoch, then, takes Kant's moral philosophy to be different in extremely important respects from the Kantianism she takes to contribute to the Prevailing Picture. For Kant the moral ideal is certainly not the mediocre objective of cottoning on to common sense and conventions, and he admits a moral reality though one that does not belong in the natural world. Murdoch even suggests at one point that the difference between her view and Kant's is an expression of different worries and different temperaments naturally driven to adopt different strategies toward what might well be thought of as a common understanding of the moral ideal (Murdoch 1970: 30). Nonetheless the strategies are different, and as I am going to argue, they are indeed opposite. In the following two sections I will analyze the difference between these strategies by criticizing two examples of an interpretive approach that tends to understate it.

4.2 TWO UNSATISFACTORY READINGS OF MURDOCHEAN LOVE

4.2.1 “Be ye therefore perfect”: the impartiality of Murdochian love

In “Love as a Moral Emotion” (1999), Velleman considers an objection to Kantian ethics. Such a framework requires an agent to act on a maxim that he can universalize, and this entails a picture of motivation that seems to be at odds with our intuitions about the operations of love. The problem was first brought up by Bernard Williams through the discussion of a case in which a man who can save only one of two people in equal peril decides to save his wife. Williams suggests that we would not expect such an agent, insofar as we are imagining him to be acting out of love, to pause and ponder whether it is in general permissible to save one’s wife in such circumstances. Such a deliberative route would be too articulate—it would involve “one thought too many”—to be recognizable as what goes on when one acts out of love (Williams 1981: 17-18). The standard reply to this objection, as Velleman reconstructs it, has been to insist that Kantianism does not require an agent to explicitly consider the question of whether the maxim she is acting on can be universalized. It only requires that, should the maxim be one that cannot be universalized, the agent would realize it and refrain from the action. Morality, according to this line of thought, does not get in the way of the operations of love with its cumbersome deliberative structure; it only keeps them in check from afar.³⁷

³⁷ Velleman mentions as proponents of this kind of reply, among others, Allison 1990, 191–98, and Herman 1993, esp. chaps. 1, 2, and 9.

Velleman is dissatisfied with this kind of reply, because it maintains that Kantian morality manages to countenance love only because it can constrain an agent's deliberation without being involved in her motivating thoughts. The reply therefore concedes that the involvement of moral considerations in deliberation would make the operations of love impossible. According to Velleman, this position, in spite of its insistence on the fact that morality does not forbid the operations of love, accepts what he takes to be the idea animating Williams's objection, namely, that between love and morality there is a "conflict in spirit" (Velleman 1999: 341). What Velleman means by this criticism is that the position underlying the standard reply is implicitly committed to the idea that Kantian morality (which is essentially impartial, i.e. requires us to attribute the same value to all individuals) must be in tension with love, which is essentially partial, i.e. requires to value certain individuals more than others. This way of understanding the difference of perspective between Kantian morality and love leaves room for the claim that, in a Kantian framework, love is an emotion that morality can tolerate, but not for the stronger claim that, in that framework, it is possible to vindicate the thought that love is a moral emotion. Velleman's goal is precisely to defend this stronger claim, and in order to do so, he turns to Murdoch's conception of love as attention.

According to Velleman, Murdoch's conception of love brings out the resemblance between the point of view of love and the point of view of Kantian morality. He attributes to Kant the identification of the motive of morality with *Achtung* or "respect", and argues that the Kantian notion of respect, like the Murdochian notion of love, names a mode of valuation-as-vision (Velleman 1999: 344). From the context of Velleman's discussion, it is clear that what he takes to

be distinctive of such “modes of valuation-as-vision” is their objective character. He emphasizes that Murdoch describes the kind of attention to individuals required by love as “impersonal,” as “an exercise of detachment,” and as “really looking” (Murdoch 1970: 64). Velleman acknowledges that love, for Murdoch, is always directed to particular individuals; but this, he argues, is not in tension with an impartial ethical outlook according to which no particular individual has more value than any other. As he puts it: “In Murdoch’s language of impersonality, detachment, realism, and justice, there is no suggestion that particularity entails partiality” (Velleman 1999: 343).

Love and respect share, according to Velleman, not only their objective character, but also their object. He summarizes his understanding of the relation between these two attitudes by saying that respect and love are “the required minimum and the optional maximum responses to the same value” (Velleman 1999: 366). Velleman argues at length that the fact that respect is for Kant, in the first instance, respect for the law is not incompatible with its being, like love, an attitude toward a person. Kantian respect, according to Velleman, is an attitude toward the idealized rational will,³⁸ which counts as a law because it constitutes a norm for the empirical will. The idealized rational will is the intelligible essence of a person and, in this sense, her true self. Velleman concludes that “[r]espect for this law is the same attitude as respect for the person; and so it can be compared to love after all” (Velleman 1999: 344).

Whether or not this is the correct way of interpreting Kant, it seems to me obvious that it does not provide, as Velleman claims, a way of seeing respect and Murdochian love as comparable in the sense of being directed to the same object. Identifying the law to which Kantian respect is directed with the idealized rational will does not fill the gap between that law and the inexhaustibly

³⁸ Later in the paper, on p. 365, Velleman spells out this ideal as “the capacity of appreciation or valuation [...] which is distinctive of self-conscious creatures like us.”

particular constellation of traits that determines the unique individuality of a person. It has been argued that it is problematic to identify the true self of a person with the idealized rational will that constitutes the norm for her empirical self (Millgram 2004: 512), but even if we accept this identification, the alleged “true self” cannot be what Murdoch indicates as the object of love. As she writes in “The Sublime and the Good,”

Kant does not tell us to respect whole particular tangled-up historical individuals, but to respect the universal reason in their breasts. In so far as we are rational and moral we are all the same, and in some mysterious sense transcendent to history. (Murdoch 1997: 215)

Murdoch is here articulating a distance between her picture of morality and Kant’s: a distance that remains despite the fact that, as she explicitly acknowledges, Kant conceives of the ideal constituting the norm for the empirical self as something that is instantiated in particular individuals. She talks of “universal reason” (which is just another way of naming such an ideal) as residing in particular individuals’ breasts, but she emphasizes that rational nature, as something that all individuals share, is not sufficient to capture their particularity. In Murdoch’s conception of the emotion that marks us as moral beings, it is exactly the particularity of individual persons (and individual realities in general) that plays the crucial role.

I will discuss this feature of Murdoch’s proposed picture of moral life in the next section. Here I want to focus on Velleman’s description of the project in pursuit of which he wants to exploit Murdoch’s conception of love. He claims that he wants to overcome the apparent tension between love and morality by “rethinking the partiality of love” (Velleman 1999: 342). But one might wonder how Murdoch’s conception of love provides any resource for such a project. As Velleman himself correctly observes, in fact, Murdoch’s emphasis on the realism and detachment

of the kind of look she identifies with love makes clear that this attitude, as she understands it, does not entail partiality.

Velleman's idea is that Murdoch's notion of love as a fair and realistic look directed at people provides a sense in which love is not partial: it does not involve the attribution to the object of love of a value that is not also attributed to everyone else. It offers, so to speak, the *pars destruens* for his project of "rethinking" the partiality of love. The *pars construens* of Velleman's proposal is independent of Murdoch's conception and consists in maintaining that love is partial, not in the sense of being *biased*, but in the sense of being *selective*. According to Velleman, respect and love have the same object because they respond to the same value, namely rational nature (which is identified with personhood), but constitute two different kinds of response to that value. Given its purely intelligible character, rational nature is something that can be grasped intellectually but is not available through the senses; for this reason, Velleman argues, it can be the immediate object of respect, but not of love. Love, according to Velleman, is a kind of response that requires the capacity to feel that our emotional defenses are unnecessary and this, in turn, requires more than the mere intellectual awareness that the individual we confront is an instantiation of rational nature. This very fact has to be manifest to our senses through the empirical features of that individual. These empirical features, however, can either disclose or hide such a value from our view, and this is why we love certain people rather than others. For this reason, while the value a person has as a person always commands respect, love is not morally required and is only possible when that very value is observably accessible to us in a way that makes us dispose of our emotional defenses. Love is selective because our capacity to see the value of persons through their empirical features is limited and because the energy and time for finding new objects of love are also resources we have in limited amounts. The main point of this proposal

is that love does *not* differ from respect in involving the attribution to the beloved of a value that differs in quality or quantity from the one that morality requires attributing to all persons. The value in question is the same, i.e. the value possessed by any person *qua* instantiation of rational nature, which is not subject to degrees; it is only the kind of access to it that changes in the two cases. Therefore love, though inherently selective, is not partial in a way that puts it in tension with the impartiality of morality.

With this outline of Velleman's proposal in place, we are now in a position to show, first, that Velleman distorts Murdoch's conception of love, and secondly, that Murdoch's conception of love constitutes a genuine alternative to the picture of morality that Velleman attributes to Kant rather than a possible extension of it.

Velleman stresses that Murdoch herself compares her notion of love to Kant's notion of *Achtung* when she speaks of love as "the extremely difficult realization that something other than oneself is real" (Murdoch 1997: 215). However, Murdoch's acknowledgment that Kant "was marvellously near to the mark" (Murdoch 1997: 216) does not imply, as Velleman assumes, that Murdoch takes Kant to have a more or less correct story about respect which can be simply supplemented with a further story about love. What Kant got right, according to Murdoch, is the intuition that morality involves the painful recognition of the existence of others. The distance that separates him from "the mark" is not a matter of incompleteness, but a failure to understand what can effect the painful recognition.

For Murdoch, it is through its "unutterable particularity" (Murdoch 1997: 215) that reality gives us the chance of achieving this recognition, insofar as we are able to attend to it. This understanding of what enables us to acknowledge the reality of others underlies what I take to be

one of the main claims of *The Sovereignty of Good*—namely, that love is not optional for morality precisely because love just *is* the capacity to attend to the particularity of reality. Murdoch writes:

the central concept of morality is “the individual” thought of as knowable by love, thought of in the light of the command, “Be ye therefore perfect.” (Murdoch 1970: 29)

This passage contains, in a very condensed form, four elements in Murdoch’s conception of love that I take Velleman to miss because of his exclusive focus on a single feature of this conception of love, that is, its *impartial* (in the sense of *unbiased*) character.

The first element is the idea that love is at the *center* of our reflection about moral life: it is not something we can make room for as an afterthought when our picture of morality, based on other resources, is already up and running. Love is not *a* moral emotion, as Velleman wants to maintain; it is *the* emotion that makes morality possible.

The second element is that love, for Murdoch, is conceptually linked to *the individual*. Murdoch thinks that love is partially identified by its proper objects, and these are identified with individual realities (Murdoch 1970: 36). This position leaves room for objects of love different from human beings; she talks for instance of love for a moral standard or for the work of an artist (Murdoch 1970: 60, 83-84). However, when Murdoch talks about love for persons, the object of love is the individual, in the sense of the particular historical human being that I have contrasted above with the idea of instantiation of rational will. According to Velleman, in criticizing Kant for being afraid of the particular, Murdoch exaggerates the importance of this contrast because she fails to properly appreciate the fact that “the object of Kantian *Achtung* can be a universal law embodied in a particular person, or [...] a particular person as embodying something universal” (Velleman 1999: 343 n. 16). But this contrast becomes extremely important if one thinks, as

Murdoch does, that it is not the awareness of the mere numerical distinctness of others, but the awareness of their qualitative differences that makes possible for the moral agent the recognition of their existence. In the Kantian framework, this recognition is achieved through the contemplation of their rational nature; but this will not do, according to Murdoch, because our rational nature is something with respect to which we are all alike. The appreciation of the specificity of the individual, in Velleman's version of the Kantian framework, is not essential to the possibility of experiencing the emotion that is constitutive of morality, i. e. respect—and even when it comes to love, this appreciation is only (and only in some cases) a means for accessing love's real object, i.e. the rational will.

The third element can be read into the command in the light of which, according to Murdoch, the individual is to be thought of: “Be ye therefore perfect.” This verse occurs at the end of The Sermon on the Mount (Matthew, 5:48), where the model of perfection is indicated in God insofar as he loves “the evil and the good,” “the righteous and the unrighteous.” Murdoch's way of characterizing the kind of perfection we have to strive for in moral life shows that her conception of love does not allow for the sort of selectivity that characterizes Velleman's. For Murdoch, love is not a kind of valuation-as-vision directed at something of moral significance that lies beyond the specific phenomenal features which make it available to our sensibility. Rather, the moral significance of the phenomenal features of individual realities lies in the bare fact that they are the constituents of the world we have to navigate as moral agents. No kind of phenomenal feature, in so far as it is real as opposed to fantasized, has any special connection with love.³⁹

³⁹ It is significant that Murdoch quotes the famous remark by Wittgenstein's “Not how the world is, but that it is, is the mystical” (TLP 6.44) in the context of a discussion of different objects of loving attention which include nature and art. About nature, Murdoch writes that what makes

Finally, the fourth element of Murdoch's conception of love that makes it incompatible with Velleman's picture is that the notion of love and the notion of individual are linked conceptually through *knowledge*. The concept that Murdoch urges us to put at the center of moral reflection is "'the individual' thought as knowable by love." Elsewhere she writes that "[l]ove is the knowledge of the individual" (Murdoch 1970: 27). For Murdoch love is not, as for Velleman, a response to an independently given object—a response one can opt for once the cognitive work is over. Love is an attitude that, though connected to action in multiple and complex ways, is also what makes its object cognitively available in the first place. My point is not, or not just, that for Murdoch love is a cognitive attitude, but rather that, in her picture of the moral life, the practical and the cognitive dimensions are interwoven in a way that does not admit for any clear separation between them.⁴⁰

A direct consequence of this aspect of Murdoch's view is that trying to achieve a proper understanding of a situation, a person, or a work of art is something that belongs to morality as much as choosing and acting do. This is part of what the already discussed famous example of M and D was meant to establish. In the next section, I will examine Bagnoli's discussion of this example and criticize her claim that, *contra* Murdoch, Kantianism can account for the phenomenon it illustrates. This, again, will give me the opportunity to draw attention to the features that distinguish Murdoch's proposed picture of the moral life from that informing Kantianism.

of natural objects such as stones and birds opportunities for moral progress is their "sheer alien pointless independent existence" (Murdoch 1970: 82-83).

⁴⁰ On this topic cf. Diamond 2010: 71-72.

4.2.2 **The “fabric of being”: Murdoch’s integrated picture of practice and cognition**

Murdoch explains her use of the notion of attention as follows:

I have used the word ‘attention,’ which I borrow from Simone Weil, to express the idea of a just and loving gaze directed upon an individual reality. I believe this to be the characteristic and proper mark of the active moral agent. (Murdoch 1970: 33)

This way of being active is something that Murdoch thinks a moral philosophy that focuses on choice and action is unable to capture. The example of M and D in *The Sovereignty of Good*, which we have already discussed in the previous chapter, is meant to illustrate, among other things, this limitation. As we have seen, according to Murdoch, we would be naturally inclined to say that M’s is a case of moral progress, but she thinks that the metaphysical background that informs contemporary moral thought, both ordinary and philosophical, makes it impossible to register it. The reason is that, largely because of Kant’s influence, this background identifies the sphere of moral activity with that of action and M’s progress does not seem to belong to that sphere.

Bagnoli argues that Murdoch is right in attributing to Kant a view according to which activity consists in action, but wrong in concluding that this entails that a Kantian moral philosophy is unable to register moral progress in a case like M’s. Bagnoli points out that action, for Kant, is not a merely physical performance. Observable performances are actions only if “they are undertaken on the basis of a reason (that is, an incentive that has been elected as a motive)” (Bagnoli 2003: 495). A Kantian perspective does not identify a moral change with a difference in outward performance and does not attribute any special significance to the question of whether the alleged change is detectable by an external observer. Kantian ethics is characterized by a first-

personal perspective and is concerned with how the agent deliberates about her conduct, not with what an observer would be able to say about it. It is essential to the Kantian conception of agency that people act for the sake of ends that affect the moral worth of their actions, and because of this, according to Bagnoli, the Kantian view can account for M's progress.

Bagnoli argues that, from the Kantian point of view, Murdoch's claim that there is no change in how M acts is inaccurate. There is a change in M's motivations, and since motivations are constitutive of the actions they occasion, there is also a change in the way she acts, even though not an observable one. Bagnoli analyzes Murdoch's example as follows:

In M's case, we have to account for her moral change as a change in the grounds that underlie her action. M's problem is how to relate to D. A change in attitude is the result of having resolved this practical issue. M's behavior after the moral change is apparently the same as it was before, but the ground for such a behavior is now different. (Bagnoli 2003: 496)⁴¹

Bagnoli is certainly right in maintaining that a Kantian understanding of moral progress does not require a change in the agent's overt behavior, and it would be wildly unfair on Murdoch's part to assume it does. The example under discussion is in fact primarily meant to challenge the Prevailing Picture rather than a more substantively Kantian framework. However, it is not obvious that the fact that Kantianism does not conceive of action in behaviorist terms suffices to show that it can accommodate M's case. As Murdoch observes, the process of transformation M goes through might well have occurred even if D had moved far away from M or had died. In that kind of scenario, M's progress would not be describable by saying that M has come to act differently

⁴¹ To avoid confusion, I have replaced the names used by Bagnoli in this passage, "Margaret" and "Daisy," with the letters originally used by Murdoch, "M" and "D."

toward D (in the sense that she has come to act toward her on different grounds than before), for the simple reason that M would not have any occasion to interact with D after the change has occurred. Still, we would feel compelled to say that M's reflection is an instance of moral progress.

Even though Bagnoli does not consider this difficulty, her analysis might be thought to be able to overcome it. For she holds not only that according to Kant the same outward behavior has a different moral worth depending on its ground, but also that the adoption of a new end is in itself a kind of action—an internal one (Bagnoli 2003: 497). Once the notion of action is broadened in this way, M's progress can be identified with the occurrence of an action even in the scenarios in which D is absent or dead. By this move Bagnoli tries to show that her Kantian analysis can account for M's progress because her framework does not take to be morally relevant only thoughts that occasion a publicly observable performance. By Murdochian lights, however, the exclusive focus on action is limiting for moral thought even apart from a behavioristic understanding of action. Bagnoli's analysis of M's progress fails to do justice to two closely related aspects of M's progress that are for Murdoch as important as its private character. The first is its temporal structure, and the second is its cognitive dimension. I will discuss them in turn.

I think it would be very hard, and not very Kantian, to try to maintain that M's progress can consist in the acquisition of the end of behaving nicely toward D on a new ground in the case in which M knows that the realization of that end is impossible—for instance, if M knows that D is dead or gone forever. For the sake of discussion, I will bracket this difficulty for Bagnoli's account. In circumstances different from this one, it is of course possible to say that, *at the end* of the process of transformation she goes through, M acquires the end of acting nicely toward D on a new ground, whether or not she ends up in fact having the opportunity to realize that end. The notion of action, though, even when broadened to include the adoption of new ends, does not

capture what happens *during* the process. The modification of M's attitude toward D, in fact, has the character of a slow and gradual transformation, not of an instantaneous act. Murdoch's contention is that the familiar kind of experience she brings to our attention through the example of M and D is not that of pausing and taking a decision at a specific moment, but that of undergoing a change that takes time. As Murdoch writes:

[I]f we consider what the work of attention is like, how continuously it goes on, and how imperceptibly it builds up structures of value round about us, we shall not be surprised that at crucial moments of choice most of the business of choosing is already over. This does not imply that we are not free, certainly not. But it implies that the exercise of our freedom is a small piecemeal business which goes on all the time and not a grandiose leaping about unimpeded at important moments. The moral life, on this view, is something that goes on continually, not something that is switched off in between the occurrence of explicit moral choices. (Murdoch 1970: 36)

Even if the Kantian understanding of what can count as moral progress defended by Bagnoli does not require a change in the agent's overt behavior, it retains the idea that moral progress proceeds always through discrete moments of choice (which sometimes are "internal actions"), and this is precisely the idea that Murdoch wants to challenge by urging us to acknowledge that the gradual and slow progression of the change in M's attitude toward D is not an inessential detail, but something constitutive of her achievement.

I said that, from a Murdochean perspective, Bagnoli's appeal to internal actions fails to do justice not only to the temporal structure of the kind of moral progress exemplified by M, but also to its cognitive dimension. This is not because Bagnoli denies any involvement of cognition in the operations of a moral agent in a situation like M's, which would be absurd, but because her analysis of what goes on with such an agent fails to depict convincingly (a) the role played by the exercises

of her cognitive powers, (b) the objects of these exercises, and (c) the complicated relation between these exercises and the determinations of her will.

A good starting point for my discussion of this issue is provided by this passage from Bagnoli, where she concedes the involvement of beliefs in M's change:

Undergoing a change can be explained as performing an internal action, and equated to adopting an end, which is expressed in a maxim. This change concerns the agent's beliefs and attitudes. Because M adopts a new end, that of appreciating D for who she is, she becomes sensitive to the facts or details that are relevant to the realization of that end. (Bagnoli 2003: 497)

Bagnoli takes the exercises of an agent's cognitive capacities (her sensitivity to certain facts or details) to enter the scene *after* the progress has occurred: their role is to make the realization of the new end possible when this has already been adopted.⁴² The conception of freedom that informs the Kantian picture defended by Bagnoli does not allow for the idea that the capacity for knowing one's circumstances, i.e. sensitivity to the facts or details that make up those circumstances, is involved not only in *realizing* one's ends but also in a distinctive way of *acquiring* ends—one which is different from choosing. This is what I meant by saying that a Murdochian approach involves a conception of the role of cognition in practical life different from the one Bagnoli attributes to Kant and herself endorses.

This conception is tightly connected to the difference between the two approaches at the level of the relation between the exercises of a subject's cognitive and active powers. As Murdoch

⁴² I think that Bagnoli here formulates ambiguously the new end she takes M to have adopted. In the context of the other formulations of this end she provides and of her general argument, we should take "appreciating D for who she is" to be equivalent to "responding to D as an instantiation of the capacity of self-legislation."

stresses, her radical criticism of a moral philosophy centered on the notion of choice does not amount to denying that our motivation and our conduct are within the control of our will, i.e., that we are free. It is rather meant to put forward a more varied representation of how the will operates, that is, not just through choices, but also—importantly—by influencing belief in the indirect way that is appropriate. We do not and cannot believe anything at will, but our will has a role in determining the quality and direction of our attention and therefore is involved also in episodes of our moral life that are not moments of choice. What goes missing in Bagnoli's analysis is exactly this interaction between will and cognition and between the active and the passive (in the sense of receptive) aspects of the agent's relation to the reality she confronts. The framework informing this analysis, in other words, does not allow for what Murdoch describes in terms of the image of the "fabric of being" (Murdoch 1970: 21).

In the last two paragraphs I have discussed the differences between Murdoch's and Bagnoli's respective understandings of the role of cognition in moral progress and of the relation between the operations of cognition and those of the will. I now turn to their different construal of the object of the cognition involved in a case of moral progress such as M's.

For Murdoch, what the active moral agent does and should do, her "characteristic and proper mark," is engaging in the effort of knowing individual realities by attending to *their phenomenal features*. Such an effort consists in reflection about and refinement of one's own concepts as well as in criticism of one's previous applications of those concepts. The example of M and D is supposed to highlight this process of improvement of our conceptual resources and of our way of employing them.

Bagnoli agrees with Murdoch that morality depends on the recognition of the reality of others, but resists the claim that in order to achieve such a recognition we need to attend to

something more than their capacity for self-legislation (Bagnoli 2003: 505). She protests that “it is not obvious that mutual recognition requires acknowledgment of differences” (Bagnoli 2003: 504). However, she also claims that Murdoch is wrong in maintaining (in the context of her criticism of Kant’s understanding of the object of respect) that as rational beings “we are all the same” (Murdoch 1959: 215). Against this criticism, she observes that the Kantian idealized rationality is in harmony with the other instantiations of itself, but harmony only requires cooperation, not uniformity (Bagnoli 2003: 504). One might wonder why Bagnoli would emphasize that regarding other human beings as units capable of self-determination is compatible with regarding them as being different from one another given that—for her—acknowledgment of differences is inessential to mutual recognition. The reason is that Bagnoli exploits this compatibility claim in order to understate the difference between Kantian respect and Murdochian love. It is in fact the claim that persons conceived as mere instantiations of rational will are, nonetheless, different from one another that enables Bagnoli to redeploy, in connection with respect, the same terminology, associated with the idea of vision, that Murdoch uses to describe what she means by loving attention. Bagnoli writes, for instance:

[Respect] requires us to be perceptive of [the] needs and qualities [of others], and capable of acting on that perception. Thus, respect (like loving attention) is a way of looking at and attending to others, not being distracted by our own selfish tendencies. In this sense, moral progress resides in not overestimating our own interests, needs, and viewpoints. As we become less self-absorbed, we grow more attentive, perceptive of and sensitive to the claims that others have on us. (Bagnoli 2003: 507)

So, for Bagnoli, respect is a response to the capacity for self-legislation, but it requires nonetheless sensitivity to the phenomenal features of other people. In order to act in accordance with the demands of respect toward the capacity of self-legislation, in fact, an agent has to be aware of the

needs and qualities of its bearers. Without such awareness, the agent could not cooperate harmoniously with other people. Bagnoli's suggestion, then, is that the distance between Kantian respect and Murdochian love is not so significant, given that both essentially involve appreciation of the phenomenal features of other people.

I think this suggestion is to be rejected. In order to cooperate with another rational being, one only needs to be sensitive to *some* of their phenomenal features, namely, those that impinge on the possible objects of one's own will. Awareness of the boundaries that the very existence of another instantiation of rational will sets for one's own—knowing one's place, as we say—does not require and, actually, has nothing to do with knowing who this person really is: it only requires one to know a limited set of this person's preferences and choices.

In order to bring out the difference between the sensitivity required by respect and the sensitivity that Murdoch calls loving attention, it will help to refer again to the example of M and D. A vulgar, undignified, noisy adult human being is still an instantiation of the capacity of self-legislation and, just as such, as deserving of Kantian respect as her more agreeable peers. M's determination to act toward D out of duty cannot depend on D's having certain phenomenal features rather than others for the simple reason that, if M's determination were so dependent, it *eo ipso* could not be the determination to act out of duty. This is why, as I said above, the only role that Bagnoli might concede to M's progress in the knowledge of D's phenomenal features is that of making it possible to realize an end she has already acquired on different grounds, namely, the end of cooperation. But this is implausible too. The kind of harmoniously cooperative behavior that respect commands was, *ex hypothesi*, already possible for M before the change she went through. In order to cooperate with D she only needed to know her preferences and decisions. She

didn't need to settle the question of whether these preferences and decisions were expressions of vulgarity or refreshing simplicity.

I am not maintaining that Bagnoli's claim that M lacked respect for D at the beginning of the story and acquired it at the end is incompatible with anything Murdoch says in her description of the example. My contention is that Bagnoli does not provide us with an analysis of M and D's case, but rather tells us another story—one for which a Kantian analysis is fitting. What I mean by this is that, if the moral progress of M were actually a matter of adopting the end of duty (i.e. coming to be motivated by respect), her achievement of a more accurate picture of D would have been completely irrelevant to it. The distorted picture of D that M had at the beginning of the story is not an obstacle for recognizing her as a rational agent, or for cooperating with her. Therefore, one cannot explain by means of an appeal to respect why having overcome that distorted picture through an effort of attention is a moral achievement for M, not even when one notices—as Bagnoli does—that acting out of respect involves some measure of sensitivity to phenomenal features.

Bagnoli's failure to acknowledge the substantive difference between respect and loving attention leads her to refuse to grant Murdoch's moral philosophy the status of a genuine alternative to the Kantian framework:

To some extent, Murdoch is aware of the similarity between loving gaze and Kant's respect [...] Nevertheless, she continues to hold that her conception of love expresses a conception of moral agency different from Kant's respect. This is what I deny. Not only do respect and loving attention work likewise and exhibit a similar phenomenology, they also rest on a common conception of moral deliberation. (Bagnoli 2002: 507)

Bagnoli, however, fails to establish any substantive commonality between respect and loving attention either at the level of the conception of moral deliberation on which they rest or at the level of their phenomenology. As I have argued, the assimilation of the two attitudes is prevented by the fact that the frameworks to which they respectively belong attribute a radically different role to choice and to the phenomenal features of historical individuals and display a radically different understanding of the relations between cognitive powers and the will.

The Murdochean framework and the Kantian one, at least in the versions that I have considered here, are, therefore, significantly different. Furthermore, if the understanding of differences I have articulated is correct, the Murdochean one is superior to this kind of alternative not just because it is able to make visible a very specific variety of moral progress, but because it is able to do so in virtue of a better understanding of the role of cognition in moral thought and of its relation to the operations of the will.

5.0 CONCLUSION

I have presented an interpretation of Murdoch's work in moral philosophy as providing (1) an analysis of the situation of contemporary moral thought as resulting from a specific philosophical elaboration of certain major historical processes, such as the rise of modern science and the marginalization of a once dominant religious understanding of the world and (2) an alternative understanding of what a proper consideration of these historical processes entails for the possibilities of moral thought.

In Chapter 1 I have argued that Murdoch's somewhat shocking claim that contemporary moral thought is dominated by a single theory amounts to an illuminating way of identifying very deep shared features in a set of apparently distinct ways of thinking that are normally conceptualized as different theories within moral philosophy and as doctrines that belong to other subfields of philosophy, such as philosophy of mind and philosophy of language. I have also argued that Murdoch's attitude toward science and logic is not characterized by indifference, but by the acknowledgment of the specificity of the conceptual apparatus that the cultivation of these disciplines has made possible to set up. This conceptual apparatus, as Murdoch points out, is shaped by the specific needs and interests that science and logic serve, and it is this very fact that makes it inadequate to set the standard of factuality in areas of our lives that are organized by different needs and interests.

The main outcome of the discussion of this part of Murdoch's analysis is that we have no intellectual obligation to take science and logic to yield the ultimate description of the world, thereby restricting the possibilities of moral thought to the variants of what I have been calling the Prevailing Picture. By rejecting the assumption that we have this obligation, Murdoch brings back

into the space of possibilities for moral thought a different kind of picture of the soul: one that makes room for the familiar ideas that the world we inhabit strikes us as informed by a structure of value and that the manifestations of our personal subjectivity are largely constituted by our responding to what we take to be more or less comprehensive glimpses of such structure with choices and actions, but also with the slow work of adjustment of our worldview through efforts of attention.

If one is free from the philosophical prejudices that constitute her main target, according to Murdoch, she would naturally recognize as appropriate standards for the assessment of the comparative merits of her proposed picture of the soul and of the Prevailing Picture not their respective standing with respect to science and logic, which is the same, but their respective success in explaining our moral experience and in constituting a resource for moral progress. It is on the basis of these criteria that Murdoch is confident that one will also recognize as more persuasive than their existentialist counterparts the conception of the world as informed by a structure of value and the conception of the self as containing much more than its capacity for choice.

The reliance on these ideas, on Murdoch's part, does not amount to a narrow-minded confidence that her reader will, as the philosophical lingo has it, "share her intuitions." One of the deepest motivations for her criticism of the Prevailing Picture is, as we have seen, that it "admits no escape into other theories" (1970: 1); and this motivation is operative in her alternative image of moral life. Neither the sort of general orientation toward life that Murdoch calls existentialist, nor the more recognizably Kantian view I have considered in Chapter 4 become inexpressible in her framework. They just acquire their proper characterizations as, precisely, two orientations toward life among many, neither of which has a primacy over its alternatives.

Murdoch's claim that the framework she offers, or rather retrieves, is superior to its alternatives, accordingly, is not an attempt to attribute to it such a primacy, i.e. the status of ultimate picture. It is something she justifies by appeal to specific ways in which it is explanatorily and morally better than these alternatives. My aim in Chapters 2, 3, and 4 has been either to apply (Chapter 2) or to bring into sharper focus (Chapters 3 and 4) Murdoch's framework by examining each of three themes: the significance of the concept of a human being in moral thought, the relation between the privacy of our inner lives and moral progress, and the role of love in moral life. I intend my examination of these themes to be a contribution to the process of assessment of Murdoch's framework that adds to our understanding of its specificity and value.

BIBLIOGRAPHY

- Anscombe G.E.M. 1981. "Modern Moral Philosophy." In *Ethics, Religion and Politics. Collected Philosophical Papers Vol. III*. Oxford: Basil Blackwell: 26-42.
- Anscombe, G.E.M. 2008. "Contraception and Chastity." In *Faith in a Hard Ground. Essays on Religion, Philosophy and Ethics*. M. Geach and L. Gormally (ed). Imprint Academic: 170-191.
- Anscombe, G.E.M. 2005. "The Dignity of the Human Being." In *Human Life, Action and Ethics*. M. Geach and L. Gormally (ed). Imprint Academic: 67-73.
- Allison, H. E. 1990. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.
- Appiah, K. A. 2008. *Experiments in Ethics*. Cambridge, Mass.: Harvard University Press.
- Blaffer Hrdy S. 2009. *Mothers and Others: The Evolutionary Origins of Mutual Understanding*. Cambridge, Mass.: Harvard University Press.
- Bagnoli, C. 2003. "Respect and Loving Attention." *Canadian Journal of Philosophy* 35(4): 483-516.
- Chuen, L. 2017. "Food, Race, and Power: Who gets to be an authority on 'ethnic' cuisines?" *Intersectional Analyst*, published online on Jan. 8, 2017: <<http://www.intersectionalanalyst.com/intersectional-analyst/2017/1/7/who-gets-to-be-an-authority-on-ethnic-cuisines>>.
- Diamond, C. 1978. "Eating Meat and Eating People." *Philosophy* 53: 465-479.
- Diamond, C. 1991. "The Importance of Being Human." *Royal Institute of Philosophy Supplement* 29: 35-62.
- Diamond, C. 2001. "Le cas du soldat nu." Transl. by Jean-Yves Mondon. *Cités* 5: 113-125.
- Diamond, C. 2010. "Murdoch the Explorer." *Philosophical Topics* 38(1): 51-85.
- Diamond, C. 2011. "Murdoch off the Map, or Taking Empiricism Back from the Empiricists." Talk delivered at The University of Montana on October 11th, 2011.
- Donatelli, P. 2012. *La vita umana in prima persona*. Roma-Bari: Laterza.
- Gopnik, A. 2006. "Headless Horseman: The Reign of Terror Revisited." *The New Yorker*, June 5.
- Hampshire, S. 1962. Disposition and Memory. In *The Journal of Psycho-Analysis*. London Vol. 43: 59-68.

- Hampshire, S. 1970. *Thought and action*. London: Chatto and Windus.
- Haque, O. S., and A. Waytz. 2011. "Why Doctors Should Be More Empathetic—But Not Too Much More." *Scientific American*, published online on April 26, 2011: <<http://www.scientificamerican.com/article.cfm?id=doctors-and-dehumanization-effect>>.
- Herman, B. 1993. *The Practice of Moral Judgment*. Cambridge, Mass.: Harvard University Press, 1993.
- Kittay, E. F. 2005. "At the Margins of Moral Personhood." *Ethics* 116(1): 100-131.
- Korsgaard, C. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Levi, P. 1959. *If This Is A Man*. Translated from the Italian by Stuart Woolf. New York: The Orion Press.
- Levi, P. 1986. Interview about *Survival in Auschwitz*, February 17, 1986. Translated from Italian into English by Ruth Fledman. *The New Republic*, <<https://newrepublic.com/article/119959/interview-primo-levi-survival-auschwitz>>.
- Madden, T. F. 2012. *Venice: A New History*. New York: Viking Penguin.
- Manne, K. 2016. "Humanism: A Critique." *Social Theory and Practice* 42(2): 389-415.
- McDowell, J. 1996. *Mind and World*. Cambridge, Mass: Harvard University Press.
- McDowell, J. 1998. "Intentionality and Interiority in Wittgenstein." In McDowell. *Mind, Value, and Reality*. Cambridge, Mass: Harvard University Press: 297-321.
- McDowell, J. 2009. "Are Meaning, Understanding, etc., Definite States? In McDowell. *The Engaged Intellect: Philosophical Essays*. Cambridge, Massachusetts: Harvard University Press: 79-95.
- Millgram, E. 2004. "Kantian Crystallization." *Ethics* 114(3): 511-513.
- Murdoch, I. 1951. "Thinking and Language." *Proceedings of the Aristotelian Society*. Supplementary Volumes, Vol. 25, *Freedom, Language, and Reality*: 25-82
- Murdoch, I. 1997. "The Sublime and the Good." In Murdoch. *Existentialists and Mystics*. London: Chatto and Windus: 205-220.
- Murdoch, I. 1997. "Existentialists and Mystics." In Murdoch. *Existentialists and Mystics*. London: Chatto and Windus: 221-234.
- Murdoch, I. 1970. *The Sovereignty of Good*. London: Routledge.
- Orwell, G. 1981. *A Collection of Essays*. New York: Harcourt.
- Singer, P. 1974. "All Animals Are Equal." *Philosophic Exchange* 5(1): 103-116.

- Smith, D. L. 2016. "Paradoxes of Dehumanization." *Social Theory and Practice* 42(2): 416-443.
- Strawson, P. F. 1962. "Freedom and Resentment," in *Proceedings of the British Academy* 48: 187-211.
- Thompson, M. 2004. "Apprehending a Human Form." *Royal Institute of Philosophy Supplement* 54: 47-74.
- Velleman, D. 1999. "Love as a Moral Emotion." *Ethics* 109(2): 338-374.
- Williams, B. 1993. *Shame and Necessity*. Berkeley and Los Angeles: University of California Press.
- Wittgenstein, L. 1922 (TLP). *Tractatus Logico-Philosophicus*. Translated by C. K. Ogden and F. Ramsey. London: Routledge and Kegan Paul.
- Wittgenstein, L. 1958 (PI). *Philosophical Investigations*. 2nd edition. Translated by G.E.M. Anscombe. Oxford: Blackwell.
- Wolf, S. 2015. "Responsibility, Moral and Otherwise." *Inquiry* 58(2): 127-142.
- Wright, C. 1989. "Wittgenstein's Later Philosophy of mind: Sensation, Privacy, and Intention." *Journal of Philosophy* 86(11): 622-634.
- Zhanqing, Y. 2016. "China: A Life in Detention." *New York Review of Books*, October 6, 2016, <<http://www.nybooks.com/contributors/yang-zhangqing>>.